

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
28 October 2004 (28.10.2004)

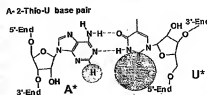
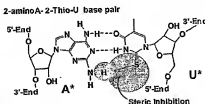
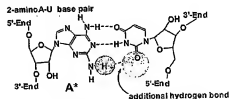
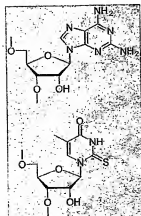
PCT

(10) International Publication Number
WO 2004/091515 A2

- (51) International Patent Classification⁷: **A61K** US04/07070 8 March 2004 (08.03.2004) US
US04/10586 5 April 2004 (05.04.2004) US
- (21) International Application Number: PCT/US2004/011255 (71) Applicant (for all designated States except US): AL-
NYLAM PHARMACEUTICALS, INC. [US/US]; 790
Memorial Dr., Suite 202, Cambridge, MA 02139 (US).
- (22) International Filing Date: 9 April 2004 (09.04.2004)
- (25) Filing Language: English (72) Inventors; and
(75) Inventors/Applicants (for US only): MANOHARAN,
Muthiah [US/US]; 25 Circle Dr., Weston, MA 02493
(US). ELBASHIR, Sayda [DE/US]; 12 Grozier Road,
Cambridge, MA 02138 (US). HARBORTH, Jens
[DE/US]; 16 Chauncy St., Apt. 35, Cambridge, MA 02138
(US).
- (26) Publication Language: English
- (30) Priority Data:
- | | | |
|------------|--------------------------------|----|
| 60/462,097 | 9 April 2003 (09.04.2003) | US |
| 60/461,915 | 10 April 2003 (10.04.2003) | US |
| 60/462,894 | 14 April 2003 (14.04.2003) | US |
| 60/463,772 | 17 April 2003 (17.04.2003) | US |
| 60/465,802 | 25 April 2003 (25.04.2003) | US |
| 60/465,665 | 25 April 2003 (25.04.2003) | US |
| 60/469,612 | 9 May 2003 (09.05.2003) | US |
| 60/493,986 | 8 August 2003 (08.08.2003) | US |
| 60/494,597 | 11 August 2003 (11.08.2003) | US |
| 60/506,341 | 26 September 2003 (26.09.2003) | US |
| 60/510,246 | 9 October 2003 (09.10.2003) | US |
| 60/510,318 | 10 October 2003 (10.10.2003) | US |
| 60/518,453 | 7 November 2003 (07.11.2003) | US |
- (74) Agent: MYERS, Louis; Fish & Richardson, P.C., 225
Franklin Street, Boston, MA 02110-2804 (US).
- (81) Designated States (unless otherwise indicated, for every
kind of national protection available): AE, AG, AL, AM,
AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN,
CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI,
GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE,
KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD,
MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG,
PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM,

[Continued on next page]

(54) Title: iRNA CONJUGATES



(57) Abstract: Therapeutic iRNA agents and methods of making and using are enclosed.



TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

- (84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK,

Published:

- without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

IRNA CONJUGATES

RELATED APPLICATIONS

The present application claims the benefit of U.S. Provisional Application No. 60/462,097, filed April 9, 2003; U.S. Provisional Application No. 60/461,915, filed April 10, 2003; U.S. Provisional Application No. 60/463,772, filed April 17, 2003; U.S. Provisional Application No. 60/465,802, filed April 25, 2003; U.S. Provisional Application No. 60/493,986, filed August 8, 2003; U.S. Provisional Application No. 60/494,597, filed August 11, 2003; U.S. Provisional Application No. 60/506,341, filed September 26, 2003; U.S. Provisional Application No. 60/518,453, filed November 7, 2003; U.S. Provisional Application No. 60/469,612, filed May 9, 2003; U.S. Provisional Application No. 60/510,246, filed October 9, 2003; U.S. Provisional Application No. 60/510,318, filed October 10, 2003; U.S. Provisional Application No. 60/465,665, filed April 25, 2003; U.S. Provisional Application No. 60/462,894, filed April 14, 2003; International Application No. PCT/US04/07070, filed March 8, 2004; and International Application No. [xxxxxx], filed April 5, 2004. The contents of these applications are hereby incorporated by reference in their entirety.

TECHNICAL FIELD

The invention relates to RNAi and related methods, e.g., methods of making and using iRNA agents. It includes methods and compositions for silencing genes expressed in the liver, and methods and compositions for directing iRNA agents to the liver.

BACKGROUND

RNA interference or "RNAi" is a term initially coined by Fire and co-workers to describe the observation that double-stranded RNA (dsRNA) can block gene expression when it is introduced into worms (Fire *et al.*, *Nature* 391:806-811, 1998). Short dsRNA directs gene-specific, post-transcriptional silencing in many organisms, including vertebrates, and has provided a new tool for studying gene function. RNAi may involve mRNA degradation.

Attorney's Docket No.: 14174-072W01

Work in this field is typified by comparatively cumbersome approaches to delivery of dsRNA to live mammals. *E.g.*, McCaffrey *et al.* (Nature 418:38-39, 2002) demonstrated the use of dsRNA to inhibit the expression of a luciferase reporter gene in mice. The dsRNAs were administered by the method of hydrodynamic tail vein injections (in addition, inhibition
5 appeared to depend on the injection of greater than 2 mg/kg dsRNA). The inventors have discovered, *inter alia*, that the unwieldy methods typical of some reported work are not needed to provide effective amounts of dsRNA to mammals and in particular not needed to provide therapeutic amounts of dsRNA to human subjects. The advantages of the current invention include practical, uncomplicated methods of administration and therapeutic applications, *e.g.*, at
10 dosages of less than 2 mg/kg.

SUMMARY

Aspects of the invention relate to compositions and methods for silencing genes expressed in the liver, *e.g.*, to treat disorders of or related to the liver. An iRNA agent composition of the invention can be one which has been modified to alter distribution in favor of
15 the liver. A composition of the invention includes an iRNA agent, *e.g.*, an iRNA agent or sRNA agent described herein.

In one aspect, the invention features a method for reducing apoB-100 levels in a subject, *e.g.*, a mammal, such as a human. The method includes administering to a subject an iRNA agent which targets apoB-100. The iRNA agent can be one described here, and can be a dsRNA that is
20 substantially identical to a region of the apoB-100 gene. The iRNA can be less than 30 nucleotides in length, *e.g.*, 21-23 nucleotides. Preferably, the iRNA is 21 nucleotides in length. In one embodiment, the iRNA is 21 nucleotides in length, and the duplex region of the iRNA is 19 nucleotides. In another embodiment, the iRNA is greater than 30 nucleotides in length.

In a preferred embodiment, the subject is treated with an iRNA agent which targets one
25 of the sequences listed in Tables 9 or 10. In a preferred embodiment it targets both sequences of a palindromic pair provided in Tables 9 or 10. The most preferred targets are listed in descending order of preferrability, in other words, the more preferred targets are listed earlier in Tables 9 or 10.

Attorney's Docket No.: 14174-072W01

In a preferred embodiment the iRNA agent will include regions, or strands, which are complementary to a pair in Tables 9 or 10. In a preferred embodiment the iRNA agent will include regions complementary to the palindromic pairs of Tables 9 or 10 as a duplex region.

In a preferred embodiment the duplex region of the iRNA agent will target a sequence
5 listed in Tables 9 or 10 but will not be perfectly complementary with the target sequence, e.g., it will not be complementary at at least 1 base pair. Preferably it will have no more than 1, 2, 3, 4, or 5 bases, in total, or per strand, which do not hybridize with the target sequence.

The iRNA agent that targets apoB-100 can be administered in an amount sufficient to reduce expression of apoB-100 mRNA. In one embodiment, the iRNA agent is administered in
10 an amount sufficient to reduce expression of apoB-100 protein (e.g., by at least 2%, 4%, 6%, 10%, 15%, 20%). Preferably, the iRNA agent does not reduce expression of apoB-48 mRNA or protein. This can be effected, e.g., by selection of an iRNA agent which specifically targets the nucleotides subject to RNA editing in the apoB-100 transcript.

The iRNA agent that targets apoB-100 can be administered to a subject, wherein the
15 subject is suffering from a disorder characterized by elevated or otherwise unwanted expression of apoB-100, elevated or otherwise unwanted levels of cholesterol, and/or dysregulation of lipid metabolism. The iRNA agent can be administered to an individual at risk for the disorder to delay onset of the disorder or a symptom of the disorder. These disorders include HDL/LDL cholesterol imbalance; dyslipidemias, e.g., familial combined hyperlipidemia (FCHL), acquired
20 hyperlipidemia; hypercholesterolemia; statin-resistant hypercholesterolemia; coronary artery disease (CAD) coronary heart disease (CHD) atherosclerosis. In one embodiment, the iRNA that targets apoB-100 is administered to a subject suffering from statin-resistant hypercholesterolemia.

The apoB-100 iRNA agent can be administered in an amount sufficient to reduce levels
25 of serum LDL-C and/or HDL-C and/or total cholesterol in a subject. For example, the iRNA is administered in an amount sufficient to decrease total cholesterol by at least 0.5%, 1%, 2.5%, 5%, 10% in the subject. In one embodiment, the iRNA agent is administered in an amount sufficient to reduce the risk of myocardial infarction the subject.

In a preferred embodiment the iRNA agent is administered repeatedly. Administration of
30 an iRNA agent can be carried out over a range of time periods. It can be administered daily, once every few days, weekly, or monthly. The timing of administration can vary from patient to

Attorney's Docket No.: 14174-072W01

patient, depending on such factors as the severity of a patient's symptoms. For example, an effective dose of an iRNA agent can be administered to a patient once a month for an indefinite period of time, or until the patient no longer requires therapy. In addition, sustained release compositions containing an iRNA agent can be used to maintain a relatively constant dosage in the patient's blood.

In one embodiment, the iRNA agent can be targeted to the liver, and apoB expression level are decreased in the liver following administration of the apoB iRNA agent. For example, the iRNA agent can be complexed with a moiety that targets the liver, e.g., an antibody or ligand that binds a receptor on the liver.

The iRNA agent, particularly an iRNA agent that targets apoB, beta-catenin or glucose-6-phosphatase RNA, can be targeted to the liver, for example by associating, e.g., conjugating the iRNA agent to a lipophilic moiety, e.g., a lipid, cholesterol, oleyl, retinyl, or cholesterol residue. Other lipophilic moieties that can be associated, e.g., conjugated with the iRNA agent include cholic acid, adamantane acetic acid, 1-pyrene butyric acid, dihydrotestosterone, 1,3-Bis-O(hexadecyl)glycerol, geranyloxyhexyl group, hexadecylglycerol, borneol, menthol, 1,3-propanediol, heptadecyl group, palmitic acid, myristic acid, O3-(oleoyl)lithocholic acid, O3-(oleoyl)cholenic acid, dimethoxytrityl, or phenoxazine. In one embodiment, the iRNA agent can be targeted to the liver by associating, e.g., conjugating, the iRNA agent to a low-density lipoprotein (LDL), e.g., a lactosylated LDL. In another embodiment, the iRNA agent can be targeted to the liver by associating, e.g., conjugating, the iRNA agent to a polymeric carrier complex with sugar residues.

In another embodiment, the iRNA agent can be targeted to the liver by associating, e.g., conjugating, the iRNA agent to a liposome complexed with sugar residues. A targeting agent that incorporates a sugar, e.g., galactose and/or analogues thereof, is particularly useful. These agents target, in particular, the parenchymal cells of the liver (see Table 1). In a preferred embodiment, the targeting moiety includes more than one galactose moieties, preferably two or three. Preferably, the targeting moiety includes 3 galactose moieties, e.g., spaced about 15 angstroms from each other. The targeting moiety can be lactose. A lactose is a glucose coupled to a galactose. Preferably, the targeting moiety includes three lactoses. The targeting moiety can also be N-Acetyl-Galactosamine, N-Ac-Glucosamine. A mannose, or mannose-6-phosphate targeting moiety can be used for macrophage targeting.

Attorney's Docket No.: 14174-072W01

The targeting agent can be linked directly, e.g., covalently or non covalently, to the iRNA agent, or to another delivery or formulation modality, e.g., a liposome. E.g., the iRNA agents with or without a targeting moiety can be incorporated into a delivery modality, e.g., a liposome, with or without a targeting moiety.

5 It is particularly preferred to use an iRNA conjugated to a lipophilic molecule to conjugate to an iRNA agent that targets apoB, beta-catenin or glucose-6-phosphatase iRNA targeting agent.

In one embodiment, the iRNA agent has been modified, or is associated with a delivery agent, e.g., a delivery agent described herein, e.g., a liposome, which has been modified to alter
10 distribution in favor of the liver. In one embodiment, the modification mediates association with a serum albumin (SA), e.g., a human serum albumin (HSA), or a fragment thereof.

The iRNA agent, particularly an iRNA agent that targets apoB, beta-catenin or glucose-6-phosphatase RNA, can be targeted to the liver, for example by associating, e.g., conjugating the iRNA agent to an SA molecule, e.g., an HSA molecule, or a fragment thereof. In one
15 embodiment, the iRNA agent or composition thereof has an affinity for an SA, e.g., HSA, which is sufficiently high such that its levels in the liver are at least 10, 20, 30, 50, or 100% greater in the presence of SA, e.g., HSA, or is such that addition of exogenous SA will increase delivery to the liver. These criteria can be measured, e.g., by testing distribution in a mouse in the presence or absence of exogenous mouse or human SA.

20 The SA, e.g., HSA, targeting agent can be linked directly, e.g., covalently or non-covalently, to the iRNA agent, or to another delivery or formulation modality, e.g., a liposome. E.g., the iRNA agents with or without a targeting moiety can be incorporated into a delivery modality, e.g., a liposome, with or without a targeting moiety.

It is particularly preferred to use an iRNA conjugated to an SA, e.g., an HSA, molecule
25 wherein the iRNA agent is an apoB, beta-catenin or glucose-6-phosphatase iRNA targeting agent.

In another aspect, the invention features, a method for reducing glucose-6-phosphatase levels in a subject, e.g., a mammal, such as a human. The method includes administering to a subject an iRNA agent which targets glucose-6-phosphatase. The iRNA agent can be a dsRNA
30 that has a sequence that is substantially identical to a sequence of the glucose-6-phosphatase gene.

Attorney's Docket No.: 14174-072W01

In a preferred embodiment, the subject is treated with an iRNA agent that targets one of the sequences listed in Table 11. In a preferred embodiment it targets both sequences of a palindromic pair provided in Table 11. The most preferred targets are listed in descending order of preferability, in other words, the more preferred targets are listed earlier in Table 11.

5 In a preferred embodiment the iRNA agent will include regions, or strands, which are complementary to a pair in Table 11. In a preferred embodiment the iRNA agent will include regions complementary to the palindromic pairs of Table 11 as a duplex region.

In a preferred embodiment the duplex region of the iRNA agent will target a sequence listed in Table 11 but will not be perfectly complementary with the target sequence, e.g., it will
10 not be complementary at at least 1 base pair. Preferably it will have no more than 1, 2, 3, 4, or 5 bases, in total, or per strand, which do not hybridize with the target sequence

In a preferred embodiment the iRNA agent includes overhangs, e.g., 3' or 5' overhangs, preferably one or more 3' overhangs. Overhangs are discussed in detail elsewhere herein but are preferably about 2 nucleotides in length. The overhangs can be complementary to the gene
15 sequences being targeted or can be other sequence. TT is a preferred overhang sequence. The first and second iRNA agent sequences can also be joined, e.g., by additional bases to form a hairpin, or by other non-base linkers.

Table 11 refers to sequences from human glucose-6-phosphatase. Table 12 refers to sequences from rat glucose-6-phosphatase. The sequences from table 12 can be used, e.g., in
20 experiments with rats or cultured rat cells.

In a preferred embodiment iRNA agent can have any architecture, e.g., architecture described herein. E.g., it can be incorporated into an iRNA agent having an overhang structure, overall length, hairpin vs. two-strand structure, as described herein. In addition, monomers other than naturally occurring ribonucleotides can be used in the selected iRNA agent.

25 The iRNA that targets glucose-6-phosphatase can be administered in an amount sufficient to reduce expression of glucose-6-phosphatase mRNA.

The iRNA that targets glucose-6-phosphatase can be administered to a subject to inhibit hepatic glucose production, for the treatment of glucose-metabolism-related disorders, such as diabetes, e.g., type-2-diabetes mellitus. The iRNA agent can be administered to an individual at
30 risk for the disorder to delay onset of the disorder or a symptom of the disorder.

Attorney's Docket No.: 14174-072W01

In other embodiments, iRNA agents having sequence similarity to the following genes can also be used to inhibit hepatic glucose production. These other genes include "forkhead homologue in rhabdomyosarcoma (FKHR); glucagon; glucagon receptor; glycogen phosphorylase; PPAR-Gamma Coactivator (PGC-1); Fructose-1,6-bisphosphatase; glucose-6-phosphate locator; glucokinase inhibitory regulatory protein; and phosphoenolpyruvate carboxykinase (PEPCK).

In one embodiment, the iRNA agent can be targeted to the liver, and RNA expression levels of the targeted genes are decreased in the liver following administration of the iRNA agent.

The iRNA agent can be one described herein, and can be a dsRNA that has a sequence that is substantially identical to a sequence of a target gene. The iRNA can be less than 30 nucleotides in length, e.g., 21-23 nucleotides. Preferably, the iRNA is 21 nucleotides in length. In one embodiment, the iRNA is 21 nucleotides in length, and the duplex region of the iRNA is 19 nucleotides. In another embodiment, the iRNA is greater than 30 nucleotides in length.

In another aspect, the invention features a method for reducing beta-catenin levels in a subject, e.g., a mammal, such as a human. The method includes administering to a subject an iRNA agent that targets beta-catenin. The iRNA agent can be one described herein, and can be a dsRNA that has a sequence that is substantially identical to a sequence of the beta-catenin gene. The iRNA can be less than 30 nucleotides in length, e.g., 21-23 nucleotides. Preferably, the iRNA is 21 nucleotides in length. In one embodiment, the iRNA is 21 nucleotides in length, and the duplex region of the iRNA is 19 nucleotides. In another embodiment, the iRNA is greater than 30 nucleotides in length.

In a preferred embodiment, the subject is treated with an iRNA agent which targets one of the sequences listed in Table 13. In a preferred embodiment it targets both sequences of a palindromic pair provided in Table 13. The most preferred targets are listed in descending order of preferability, in other words, the more preferred targets are listed earlier in Table 13.

In a preferred embodiment the iRNA agent will include regions, or strands, which are complementary to a pair in Table 13. In a preferred embodiment the iRNA agent will include regions complementary to the palindromic pairs of Table 13 as a duplex region.

In a preferred embodiment the duplex region of the iRNA agent will target a sequence listed in Table 13 but will not be perfectly complementary with the target sequence, e.g., it will

Attorney's Docket No.: 14174-072W01

not be complementary at at least 1 base pair. Preferably it will have no more than 1, 2, 3, 4, or 5 bases, in total, or per strand, which do not hybridize with the target sequence

In a preferred embodiment the iRNA agent includes overhangs, e.g., 3' or 5' overhangs, preferably one or more 3' overhangs. Overhangs are discussed in detail elsewhere herein but are preferably about 2 nucleotides in length. The overhangs can be complementary to the gene sequences being targeted or can be other sequence. TT is a preferred overhang sequence. The first and second iRNA agent sequences can also be joined, e.g., by additional bases to form a hairpin, or by other non-base linkers.

The iRNA agent that targets beta-catenin can be administered in an amount sufficient to reduce expression of beta-catenin mRNA. In one embodiment, the iRNA agent is administered in an amount sufficient to reduce expression of beta-catenin protein (e.g., by at least 2%, 4%, 6%, 10%, 15%, 20%).

The iRNA agent that targets beta-catenin can be administered to a subject, wherein the subject is suffering from a disorder characterized by unwanted cellular proliferation in the liver or of liver tissue, e.g., metastatic tissue originating from the liver. Examples include, a benign or malignant disorder, e.g., a cancer, e.g., a hepatocellular carcinoma (HCC), hepatic metastasis, or hepatoblastoma.

The iRNA agent can be administered to an individual at risk for the disorder to delay onset of the disorder or a symptom of the disorder

In a preferred embodiment the iRNA agent is administered repeatedly. Administration of an iRNA agent can be carried out over a range of time periods. It can be administered daily, once every few days, weekly, or monthly. The timing of administration can vary from patient to patient, depending on such factors as the severity of a patient's symptoms. For example, an effective dose of an iRNA agent can be administered to a patient once a month for an indefinite period of time, or until the patient no longer requires therapy. In addition, sustained release compositions containing an iRNA agent can be used to maintain a relatively constant dosage in the patient's blood.

In one embodiment, the iRNA agent can be targeted to the liver, and beta-catenin expression level are decreased in the liver following administration of the beta-catenin iRNA agent. For example, the iRNA agent can be complexed with a moiety that targets the liver, e.g., an antibody or ligand that binds a receptor on the liver.

Attorney's Docket No.: 14174-072W01

In another aspect, the invention provides methods to treat liver disorders, *e.g.*, disorders characterized by unwanted cell proliferation, hematological disorders, disorders characterized by inflammation disorders, and metabolic or viral diseases or disorders of the liver. A proliferation disorder of the liver can be, for example, a benign or malignant disorder, *e.g.*, a cancer, *e.g.*, a hepatocellular carcinoma (HCC), hepatic metastasis, or hepatoblastoma. A hepatic hematology or inflammation disorder can be a disorder involving clotting factors, a complement-mediated inflammation or a fibrosis, for example. Metabolic diseases of the liver can include dyslipidemias, and irregularities in glucose regulation. Viral diseases of the liver can include hepatitis C or hepatitis B. In one embodiment, a liver disorder is treated by administering one or more iRNA agents that have a sequence that is substantially identical to a sequence in a gene involved in the liver disorder.

In one embodiment an iRNA agent to treat a liver disorder has a sequence which is substantially identical to a sequence of the beta-catenin or c-jun gene. In another embodiment, such as for the treatment of hepatitis C or hepatitis B, the iRNA agent can have a sequence that is substantially identical to a sequence of a gene of the hepatitis C virus or the hepatitis B virus, respectively. For example, the iRNA agent can target the 5' core region of HCV. This region lies just downstream of the ribosomal toe-print straddling the initiator methionine. Alternatively, an iRNA agent of the invention can target any one of the nonstructural proteins of HCV: NS3, 4A, 4B, 5A, or 5B. For the treatment of hepatitis B, an iRNA agent can target the protein X (HBx) gene, for example.

In a preferred embodiment, the subject is treated with an iRNA agent which targets one of the sequences listed in Table 14. In a preferred embodiment it targets both sequences of a palindromic pair provided in Table 14. The most preferred targets are listed in descending order of preferability, in other words, the more preferred targets are listed earlier in Table 14.

In a preferred embodiment the iRNA agent will include regions, or strands, which are complementary to a pair in Table 14. In a preferred embodiment the iRNA agent will include regions complementary to the palindromic pairs of Table 14 as a duplex region.

In a preferred embodiment the duplex region of the iRNA agent will target a sequence listed in Table 14, but will not be perfectly complementary with the target sequence, *e.g.*, it will not be complementary at at least 1 base pair. Preferably it will have no more than 1, 2, 3, 4, or 5 bases, in total, or per strand, which do not hybridize with the target sequence

Attorney's Docket No.: 14174-072W01

In a preferred embodiment the iRNA agent includes overhangs, e.g., 3' or 5' overhangs, preferably one or more 3' overhangs. Overhangs are discussed in detail elsewhere herein but are preferably about 2 nucleotides in length. The overhangs can be complementary to the gene sequences being targeted or can be other sequence. TT is a preferred overhang sequence. The first and second iRNA agent sequences can also be joined, e.g., by additional bases to form a hairpin, or by other non-base linkers.

In another aspect, an iRNA agent can be administered to modulate blood clotting, e.g., to reduce the tendency to form a blood clot. In a preferred embodiment the iRNA agent targets Factor V expression, preferably in the liver. One or more iRNA agents can be used to target a wild type allele, a mutant allele, e.g., the Leiden Factor V allele, or both. Such administration can be used to treat or prevent venous thrombosis, e.g., deep vein thrombosis or pulmonary embolism, or another disorder caused by elevated or otherwise unwanted expression of Factor V, in, e.g., the liver. In one embodiment the iRNA agent can treat a subject, e.g., a human who has Factor V Leiden or other genetic trait associated with an unwanted tendency to form blood clots.

In a preferred embodiment administration of an iRNA agent which targets Factor V is with the administration of a second treatment, e.g. a treatment which reduces the tendency of the blood to clot, e.g., the administration of heparin or of a low molecular weight heparin.

In one embodiment, the iRNA agent that targets Factor V can be used as a prophylaxis in patients, e.g., patients with Factor V Leiden, who are placed at risk for a thrombosis, e.g., those about to undergo surgery, in particular those about to undergo high-risk surgical procedures known to be associated with formation of venous thrombosis, those about to undergo a prolonged period of relative inactivity, e.g., on a motor vehicle, train or airplane flight, e.g., a flight or other trip lasting more than three or five hours. Such a treatment can be an adjunct to the therapeutic use of low molecular weight (LMW) heparin prophylaxis.

In another embodiment, the iRNA agent that targets Factor V can be administered to patients with Factor V Leiden to treat deep vein thrombosis (DVT) or pulmonary embolism (PE). Such a treatment can be an adjunct to (or can replace) therapeutic uses of heparin or coumadin. The treatment can be administered by inhalation or generally by pulmonary routes.

In a preferred embodiment, an iRNA agent administered to treat a liver disorder is targeted to the liver. For example, the iRNA agent can be complexed with a targeting moiety, e.g., an antibody or ligand that recognizes a liver-specific receptor.

Attorney's Docket No.: 14174-072W01

The invention also includes preparations, including substantially pure or pharmaceutically acceptable preparations of iRNA agents which silence any of the genes discussed herein and in particular for any of apoB-100, glucose-6-phosphatase, beta-catenin, factor V, or any of the HVC genes discussed herein.

5 The methods and compositions of the invention, *e.g.*, the methods and compositions to treat diseases and disorders of the liver described herein, can be used with any of the iRNA agents described. In addition, the methods and compositions of the invention can be used for the treatment of any disease or disorder described herein, and for the treatment of any subject, *e.g.*, any animal, any mammal, such as any human.

10 The methods and compositions of the invention, *e.g.*, the methods and iRNA compositions to treat liver-based diseases described herein, can be used with any dosage and/or formulation described herein, as well as with any route of administration described herein.

A "substantially identical" sequence includes a region of sufficient homology to the target gene, and is of sufficient length in terms of nucleotides, that the iRNA agent, or a fragment thereof, can mediate down regulation of the target gene. Thus, the iRNA agent is or includes a region which is at least partially, and in some embodiments fully, complementary to a target RNA transcript. It is not necessary that there be perfect complementarity between the iRNA agent and the target, but the correspondence must be sufficient to enable the iRNA agent, or a cleavage product thereof, to direct sequence specific silencing, *e.g.*, by RNAi cleavage of the target RNA, *e.g.*, mRNA. Complementarity, or degree of homology with the target strand, is most critical in the antisense strand. While perfect complementarity, particularly in the antisense strand, is often desired some embodiments can include, particularly in the antisense strand, one or more but preferably 6, 5, 4, 3, 2, or fewer mismatches (with respect to the target RNA). The mismatches, particularly in the antisense strand, are most tolerated in the terminal regions and if present are preferably in a terminal region or regions, *e.g.*, within 6, 5, 4, or 3 nucleotides of the 5' and/or 3' terminus. The sense strand need only be sufficiently complementary with the antisense strand to maintain the over all double strand character of the molecule.

20 The details of one or more embodiments of the invention are set forth in the accompanying drawings and the description below. Other features, objects, and advantages of the invention will be apparent from this description, and from the claims. This application

30

Attorney's Docket No.: 14174-072W01

incorporates all cited references, patents, and patent applications by references in their entirety for all purposes.

BRIEF DESCRIPTION OF THE DRAWINGS

5 FIG. 1 is a structural representation of base pairing in pseudocomplementary siRNA².

FIG. 2 is a schematic representation of dual targeting siRNAs designed to target the HCV genome.

FIG. 3 is a schematic representation of pseudocomplementary, bifunctional siRNAs designed to target the HCV genome.

10 FIG. 4 is a general synthetic scheme for incorporation of RRMS monomers into an oligonucleotide.

FIG. 5 is a table of representative RRMS carriers. Panel 1 shows pyrroline-based RRMSs; panel 2 shows 3-hydroxyproline-based RRMSs; panel 3 shows piperidine-based RRMSs; panel 4 shows morpholine and piperazine-based RRMSs; and panel 5 shows decalin-based RRMSs. R1 is succinate or phosphoramidate and R2 is H or a conjugate ligand.

15 FIG. 6A is a graph depicting blood glucose levels in mice treated with nonspecific Renilla RNA or not treated with siRNA. Mice treated with nonspecific Renilla RNA were injected on Day 7.

20 FIG. 6B is a graph depicting blood glucose levels in mice treated with siRNA targeting glucose 6-phosphatase. Mice treated with siRNA targeting glucose 6-phosphatase were injected on Day 7.

FIG. 6C is a graph depicting blood glucose levels in mice that were either not injected with siRNA, or were injected but the injection failed. Mice that were injected, were injected on Day 7.

25 FIG. 7 is a graph depicting average blood glucose levels in four mice treated with siRNA targeting glucose 6-phosphatase, and in four mice either treated with nonspecific Renilla RNA or not treated with siRNA (triangles). siRNA or Renilla RNA was administered on day 7 by hydrodynamic tail vein injection.

30 FIG. 8A is a graph depicting levels of luciferase mRNA in livers of CMV-Luc mice (Xanogen) following intravenous injection (iv) of buffer or siRNA into the tail vein. Each bar represents data from one mouse. RNA levels were quantified by QuantiGene Assay

Attorney's Docket No.: 14174-072W01

(Genospectra, Inc.; Fremont, CA)). The Y axis represents chemiluminescence values in counts per second (CPS).

FIG. 8B is a graph depicting levels of luciferase mRNA in livers of CMV-Luc mice (Xanogen). The values are averaged from the data depicted in FIG. 8A.

FIG 9 is a graph depicting the pharmacokinetics of cholesterol-conjugated and unconjugated siRNA. The diamonds represent the amount of unconjugated ³³P-labeled siRNA (ALN-3000) in mouse plasma over time; the squares represent the amount of cholesterol-conjugated ³³P-labeled siRNA (ALN-3001) in mouse plasma over time. "L1163" is equivalent to ALN3000; "L1163Chol" is equivalent to ALN-3001.

FIG. 10 is a graph indicating the amount of cholesterol-conjugated (dark bars) and unconjugated siRNA (light bars) detected in mouse whole liver tissue isolated over a period of time following intravenous tail vein injection. The amount of siRNA is represented as a percentage of the total dose or ³³P-labeled siRNA delivered to the mouse. "L1163" is equivalent to ALN3000 (light bars); "L1163Chol" is equivalent to ALN-3001 (dark bars).

FIG. 11 is a graph indicating the amount of cholesterol-conjugated siRNA detected in various tissues of two different CMV-Luc mice ("Mouse 69" (light bars) and "Mouse 63" (dark bars)). Mice were injected with 50 mg/kg AL-3001 siRNA by intravenous tail vein injection, and tissue was harvested 22 hours later. siRNA was detected by RNase protection, and phosphorimager scanning was used to quantitate the siRNA. The amount of siRNA is expressed as ug/g liver tissue.

FIG. 12 is a gel of U/U siRNA (see Table 19) detected in the liver of Balbc mice at increasing time points following hydrodynamic (hd) tail vein injection. U/U siRNA was injected at a concentration of 4 mg/kg. siRNA was detected by RNase protection assay. Lanes labeled "stand." were loaded with clean siRNA to serve as size and quality standards. "non" represents control samples isolated from livers of mice that were not injected with U/U siRNA. The control samples were further used in parallel RNase protection assays.

FIG. 13 is a gel comparing different siRNA species detected in the livers of Balbc mice at increasing time points following hydrodynamic (hd) or nonhydrodynamic (iv) tail vein injection. U/U siRNA was injected by hd and by iv injection. 3°C/3°C and 3°C/U (see Table 19) were each injected by iv injection. at a concentration of 4 mg/kg. siRNA was detected by RNase protection assay. Lanes labeled "stand." were loaded with clean siRNA to serve as size and

Attorney's Docket No.: 14174-072W01

quality standards. "non" represents control samples isolated from livers of mice that were not injected with siRNA. The control samples were further used in parallel RNase protection assays.

FIG 14 is a graph depicting the percentage of luciferase activity in liver extracts of CMV-
5 Luc mice injected with siRNA (ALN-3001). Percentage of luciferase activity was relative to activity in CMV-Luc mice injected with PBS, pH 4.7. "Buffer1 siRNA1," "Buffer2 siRNA2," and "Buffer3 siRNA3" represent the average activity observed in three separate experiments.

DETAILED DESCRIPTION

10 Double-stranded (dsRNA) directs the sequence-specific silencing of mRNA through a process known as RNA interference (RNAi). The process occurs in a wide variety of organisms, including mammals and other vertebrates.

It has been demonstrated that 21-23 nt fragments of dsRNA are sequence-specific mediators of RNA silencing, *e.g.*, by causing RNA degradation. While not wishing to be bound
15 by theory, it may be that a molecular signal, which may be merely the specific length of the fragments, present in these 21-23 nt fragments recruits cellular factors that mediate RNAi. Described herein are methods for preparing and administering these 21-23 nt fragments, and other iRNAs agents, and their use for specifically inactivating gene function. The use of iRNAs agents (or recombinantly produced or chemically synthesized oligonucleotides of the same or
20 similar nature) enables the targeting of specific mRNAs for silencing in mammalian cells. In addition, longer dsRNA agent fragments can also be used, *e.g.*, as described below.

Although, in mammalian cells, long dsRNAs can induce the interferon response which is frequently deleterious, sRNAs do not trigger the interferon response, at least not to an extent that is deleterious to the cell and host. In particular, the length of the iRNA agent strands in an sRNA
25 agent can be less than 31, 30, 28, 25, or 23 nt, *e.g.*, sufficiently short to avoid inducing a deleterious interferon response. Thus, the administration of a composition of sRNA agent (*e.g.*, formulated as described herein) to a mammalian cell can be used to silence expression of a target gene while circumventing the interferon response. Further, use of a discrete species of iRNA

Attorney's Docket No.: 14174-072W01

agent can be used to selectively target one allele of a target gene, *e.g.*, in a subject heterozygous for the allele.

Moreover, in one embodiment, a mammalian cell is treated with an iRNA agent that disrupts a component of the interferon response, *e.g.*, double stranded RNA (dsRNA)-activated
5 protein kinase PKR. Such a cell can be treated with a second iRNA agent that includes a sequence complementary to a target RNA and that has a length that might otherwise trigger the interferon response.

In a typical embodiment, the subject is a mammal such as a cow, horse, mouse, rat, dog, pig, goat, or a primate. The subject can be a dairy mammal (*e.g.*, a cow, or goat) or other farmed
10 animal (*e.g.*, a chicken, turkey, sheep, pig, fish, shrimp). In a much preferred embodiment, the subject is a human, *e.g.*, a normal individual or an individual that has, is diagnosed with, or is predicted to have a disease or disorder.

Further, because iRNA agent mediated silencing persists for several days after administering the iRNA agent composition, in many instances, it is possible to administer the
15 composition with a frequency of less than once per day, or, for some instances, only once for the entire therapeutic regimen. For example, treatment of some cancer cells may be mediated by a single bolus administration, whereas a chronic viral infection may require regular administration, *e.g.*, once per week or once per month.

A number of exemplary routes of delivery are described that can be used to administer an
20 iRNA agent to a subject. In addition, the iRNA agent can be formulated according to an exemplary method described herein.

Liver Diseases

Exemplary diseases and disorders that can be treated by the methods and compositions of the invention are liver-based diseases.

25 Disorders involving the liver include, but are not limited to, hepatic injury; jaundice and cholestasis, such as bilirubin and bile formation; hepatic failure and cirrhosis, such as cirrhosis, portal hypertension, including ascites, portosystemic shunts, and splenomegaly; infectious disorders, such as viral hepatitis, including hepatitis A-E infection and infection by other

Attorney's Docket No.: 14174-072W01

hepatitis viruses, clinicopathologic syndromes, such as the carrier state, asymptomatic infection, acute viral hepatitis, chronic viral hepatitis, and fulminant hepatitis; autoimmune hepatitis; drug- and toxin-induced liver disease, such as alcoholic liver disease; inborn errors of metabolism and pediatric liver disease, such as hemochromatosis, Wilson disease, $\alpha 1$ -antitrypsin deficiency, and neonatal hepatitis; intrahepatic biliary tract disease, such as secondary biliary cirrhosis, primary biliary cirrhosis, primary sclerosing cholangitis, and anomalies of the biliary tree; circulatory disorders, such as impaired blood flow into the liver, including hepatic artery compromise and portal vein obstruction and thrombosis, impaired blood flow through the liver, including passive congestion and centrilobular necrosis and peliosis hepatis, hepatic vein outflow obstruction, including hepatic vein thrombosis (Budd-Chiari syndrome) and veno-occlusive disease; hepatic disease associated with pregnancy, such as preeclampsia and eclampsia, acute fatty liver of pregnancy, and intrahepatic cholestasis of pregnancy; hepatic complications of organ or bone marrow transplantation, such as drug toxicity after bone marrow transplantation, graft-versus-host disease and liver rejection, and nonimmunologic damage to liver allografts; tumors and tumorous conditions, such as nodular hyperplasias, adenomas, and malignant tumors, including primary carcinoma of the liver and metastatic tumors.

An iRNA agent can also be administered to inhibit Factor V expression in the liver. Two to five percent of the United States population is heterozygous for an allele of the Factor V gene that encodes a single amino acid change at position 1961. These heterozygous individuals have a 3-8 fold increased risk of venous thrombosis, a risk that is associated with increased factor V activity. The increased activity leads to increased thrombin generation from the prothrombinase complex. An iRNA agent directed against Factor V can treat or prevent venous thrombosis or treat a human who has Factor V Leiden. The iRNA agent that targets Factor V can be also be used as a prophylaxis in patients with Factor V Leiden who undergo high-risk surgical procedures, and this prophylaxis can be an adjunct to the therapeutic use of low molecular weight (LMW) heparin prophylaxis.

An iRNA agent that targets Factor V can also be administered to patients with Factor V Leiden to treat deep vein thrombosis (DVT) or pulmonary embolism (PE), and this treatment can be an adjunct to therapeutic uses of heparin or coumadin. Any other disorder caused by elevated or otherwise unwanted levels of Factor V protein can be treated by administering an iRNA agent against Factor V.

Attorney's Docket No.: 14174-072W01

iRNA agents of the invention can be targeted to any gene whose overexpression is associated with the liver diseases.

Targeting to the Liver

5 The iRNA agents of the invention are particularly useful when targeted to the liver. An iRNA agent can be targeted to the liver through a composition that includes the iRNA agent and a liver-targeting agent. For example, a liver-targeting agent can be a lipophilic moiety. Preferred lipophilic moieties include lipid, cholesterol, oleyl, retinyl, or cholesterol residues (see Table 1). Other lipophilic moieties that can function as liver-targeting agents include cholic acid, 10 adamantane acetic acid, 1-pyrene butyric acid, dihydrotestosterone, 1,3-Bis-O(hexadecyl)glycerol, geranyloxyhexyl group, hexadecylglycerol, borneol, menthol, 1,3-propanediol, heptadecyl group, palmitic acid, myristic acid, O3-(oleoyl)lithocholic acid, O3-(oleoyl)cholenic acid, dimethoxytrityl, or phenoxazine.

15 An iRNA agent can also be targeted to the liver by association with a low-density lipoprotein (LDL), such as lactosylated LDL. Polymeric carriers complexed with sugar residues can also function to target iRNA agents to the liver.

A targeting agent that incorporates a sugar, e.g., galactose and/or analogues thereof, is particularly useful. These agents target, in particular, the parenchymal cells of the liver (see Table 1). For example, a targeting moiety can include more than one or preferably two or three 20 galactose moieties, spaced about 15 angstroms from each other. The targeting moiety can alternatively be lactose (e.g., three lactose moieties), which is glucose coupled to a galactose. The targeting moiety can also be N-Acetyl-Galactosamine, N-Ac-Glucosamine. A mannose or mannose-6-phosphate targeting moiety can be used for macrophage targeting.

25 Conjugation of an iRNA agent with a serum albumin (SA), such as human serum albumin, can also be used to target the iRNA agent to the liver.

An iRNA agent can be targeted to a particular cell type in the liver by using specific targeting agents, which recognize particular receptors in the liver. Exemplary targeting moieties and their associated receptors are presented in Table 1.

Table 1 Targeting agents (Ligands) and their associated receptors

<u>Liver Cells</u>	<u>Ligand</u>	<u>Receptor</u>
1) Parenchymal Cell (PC) (Hepatocytes)	Galactose	ASGP-R (Asialoglycoprotein receptor)
	Gal NAc (n-acetyl-galactosamine)	ASPG-R Gal NAc Receptor
	Lactose	
	Asialofetuin	ASPG-r
2) Sinusoidal Endothelial Cell (SEC)	Hyaluronan	Hyaluronan receptor
	Procollagen	Procollagen receptor
	Negatively charged molecules	Scavenger receptors
	Mannose	Mannose receptors
	N-acetyl Glucosamine	Scavenger receptors
	Immunoglobulins	Fc Receptor
	LPS	CD14 Receptor
	Insulin	Receptor mediated transcytosis
	Transferrin	Receptor mediated transcytosis
	Albumins	Non-specific
	Sugar-Albumin conjugates	
3) Kupffer Cell (KC)	Mannose-6-phosphate	Mannose-6-phosphate receptor
	Mannose	Mannose receptors
	Fucose	Fucose receptors
	Albumins	Non-specific
	Mannose-albumin conjugates	

5 IRNA AGENT STRUCTURE

Described herein are isolated iRNA agents, e.g., RNA molecules, (double-stranded; single-stranded) that mediate RNAi. The iRNA agents preferably mediate RNAi with respect to an endogenous gene of a subject or to a gene of a pathogen.

An "RNA agent" as used herein, is an unmodified RNA, modified RNA, or nucleoside surrogate, all of which are defined herein (see, e.g., the section below entitled RNA Agents). While numerous modified RNAs and nucleoside surrogates are described, preferred examples

Attorney's Docket No.: 14174-072W01

include those which have greater resistance to nuclease degradation than do unmodified RNAs. Preferred examples include those which have a 2' sugar modification, a modification in a single strand overhang, preferably a 3' single strand overhang, or, particularly if single stranded, a 5' modification which includes one or more phosphate groups or one or more analogs of a phosphate group.

An "iRNA agent" as used herein, is an RNA agent which can, or which can be cleaved into an RNA agent which can, down regulate the expression of a target gene, preferably an endogenous or pathogen target RNA. While not wishing to be bound by theory, an iRNA agent may act by one or more of a number of mechanisms, including post-transcriptional cleavage of a target mRNA sometimes referred to in the art as RNAi, or pre-transcriptional or pre-translational mechanisms. An iRNA agent can include a single strand or can include more than one strands, *e.g.*, it can be a double stranded iRNA agent. If the iRNA agent is a single strand it is particularly preferred that it include a 5' modification which includes one or more phosphate groups or one or more analogs of a phosphate group.

The iRNA agent should include a region of sufficient homology to the target gene, and be of sufficient length in terms of nucleotides, such that the iRNA agent, or a fragment thereof, can mediate down regulation of the target gene. (For ease of exposition the term nucleotide or ribonucleotide is sometimes used herein in reference to one or more monomeric subunits of an RNA agent. It will be understood herein that the usage of the term "ribonucleotide" or "nucleotide", herein can, in the case of a modified RNA or nucleotide surrogate, also refer to a modified nucleotide, or surrogate replacement moiety at one or more positions.) Thus, the iRNA agent is or includes a region which is at least partially, and in some embodiments fully, complementary to the target RNA. It is not necessary that there be perfect complementarity between the iRNA agent and the target, but the correspondence must be sufficient to enable the iRNA agent, or a cleavage product thereof, to direct sequence specific silencing, *e.g.*, by RNAi cleavage of the target RNA, *e.g.*, mRNA.

Complementarity, or degree of homology with the target strand, is most critical in the antisense strand. While perfect complementarity, particularly in the antisense strand, is often desired some embodiments can include, particularly in the antisense strand, one or more but

Attorney's Docket No.: 14174-072/W01

preferably 6, 5, 4, 3, 2, or fewer mismatches (with respect to the target RNA). The mismatches, particularly in the antisense strand, are most tolerated in the terminal regions and if present are preferably in a terminal region or regions, *e.g.*, within 6, 5, 4, or 3 nucleotides of the 5' and/or 3' terminus. The sense strand need only be sufficiently complementary with the antisense strand to maintain the over all double strand character of the molecule.

As discussed elsewhere herein, an iRNA agent will often be modified or include nucleoside surrogates in addition to the RRMS. Single stranded regions of an iRNA agent will often be modified or include nucleoside surrogates, *e.g.*, the unpaired region or regions of a hairpin structure, *e.g.*, a region which links two complementary regions, can have modifications or nucleoside surrogates. Modification to stabilize one or more 3'- or 5'-terminus of an iRNA agent, *e.g.*, against exonucleases, or to favor the antisense sRNA agent to enter into RISC are also favored. Modifications can include C3 (or C6, C7, C12) amino linkers, thiol linkers, carboxyl linkers, non-nucleotidic spacers (C3, C6, C9, C12, abasic, triethylene glycol, hexaethylene glycol), special biotin or fluorescein reagents that come as phosphoramidites and that have another DMT-protected hydroxyl group, allowing multiple couplings during RNA synthesis.

iRNA agents include: molecules that are long enough to trigger the interferon response (which can be cleaved by Dicer (Bernstein *et al.* 2001. *Nature*, 409:363-366) and enter a RISC (RNAi-induced silencing complex)); and, molecules which are sufficiently short that they do not trigger the interferon response (which molecules can also be cleaved by Dicer and/or enter a RISC), *e.g.*, molecules which are of a size which allows entry into a RISC, *e.g.*, molecules which resemble Dicer-cleavage products. Molecules that are short enough that they do not trigger an interferon response are termed sRNA agents or shorter iRNA agents herein. "sRNA agent or shorter iRNA agent" as used herein, refers to an iRNA agent, *e.g.*, a double stranded RNA agent or single strand agent, that is sufficiently short that it does not induce a deleterious interferon response in a human cell, *e.g.*, it has a duplexed region of less than 60 but preferably less than 50, 40, or 30 nucleotide pairs. The sRNA agent, or a cleavage product thereof, can down regulate a target gene, *e.g.*, by inducing RNAi with respect to a target RNA, preferably an endogenous or pathogen target RNA.

Attorney's Docket No.: 14174-072W01

Each strand of an sRNA agent can be equal to or less than 30, 25, 24, 23, 22, 21, or 20 nucleotides in length. The strand is preferably at least 19 nucleotides in length. For example, each strand can be between 21 and 25 nucleotides in length. Preferred sRNA agents have a duplex region of 17, 18, 19, 29, 21, 22, 23, 24, or 25 nucleotide pairs, and one or more
5 overhangs, preferably one or two 3' overhangs, of 2-3 nucleotides.

In addition to homology to target RNA and the ability to down regulate a target gene, an iRNA agent will preferably have one or more of the following properties:

- (1) it will be of the Formula 1, 2, 3, or 4 set out in the RNA Agent section below;
- (2) if single stranded it will have a 5' modification which includes one or more
10 phosphate groups or one or more analogs of a phosphate group;
- (3) it will, despite modifications, even to a very large number, or all of the nucleosides, have an antisense strand that can present bases (or modified bases) in the proper three dimensional framework so as to be able to form correct base pairing and form a duplex structure with a homologous target RNA which is sufficient to allow down regulation of the target, e.g., by
15 cleavage of the target RNA;
- (4) it will, despite modifications, even to a very large number, or all of the nucleosides, still have "RNA-like" properties, *i.e.*, it will possess the overall structural, chemical and physical properties of an RNA molecule, even though not exclusively, or even partly, of ribonucleotide-based content. For example, an iRNA agent can contain, e.g., a sense and/or an antisense strand
20 in which all of the nucleotide sugars contain e.g., 2' fluoro in place of 2' hydroxyl. This deoxyribonucleotide-containing agent can still be expected to exhibit RNA-like properties. While not wishing to be bound by theory, the electronegative fluorine prefers an axial orientation when attached to the C2' position of ribose. This spatial preference of fluorine can, in turn, force the sugars to adopt a C_{3'}-*endo* pucker. This is the same puckering mode as
25 observed in RNA molecules and gives rise to the RNA-characteristic A-family-type helix. Further, since fluorine is a good hydrogen bond acceptor, it can participate in the same hydrogen bonding interactions with water molecules that are known to stabilize RNA structures. (Generally, it is preferred that a modified moiety at the 2' sugar position will be able to enter into

Attorney's Docket No.: 14174-072W01

H-bonding which is more characteristic of the OH moiety of a ribonucleotide than the H moiety of a deoxyribonucleotide. A preferred iRNA agent will: exhibit a C_{3'}-endo pucker in all, or at least 50, 75, 80, 85, 90, or 95 % of its sugars; exhibit a C_{3'}-endo pucker in a sufficient amount of its sugars that it can give rise to a the RNA-characteristic A-family-type helix; will have no more than 20, 10, 5, 4, 3, 2, or 1 sugar which is not a C_{3'}-endo pucker structure. These limitations are particularly preferably in the antisense strand;

(5) regardless of the nature of the modification, and even though the RNA agent can contain deoxynucleotides or modified deoxynucleotides, particularly in overhang or other single strand regions, it is preferred that DNA molecules, or any molecule in which more than 50, 60, or 70 % of the nucleotides in the molecule, or more than 50, 60, or 70 % of the nucleotides in a duplexed region are deoxyribonucleotides, or modified deoxyribonucleotides which are deoxy at the 2' position, are excluded from the definition of RNA agent.

A "single strand iRNA agent" as used herein, is an iRNA agent which is made up of a single molecule. It may include a duplexed region, formed by intra-strand pairing, e.g., it may be, or include, a hairpin or pan-handle structure. Single strand iRNA agents are preferably antisense with regard to the target molecule. In preferred embodiments single strand iRNA agents are 5' phosphorylated or include a phosphoryl analog at the 5' prime terminus. 5'-phosphate modifications include those which are compatible with RISC mediated gene silencing. Suitable modifications include: 5'-monophosphate ((HO)2(O)P-O-5'); 5'-diphosphate ((HO)2(O)P-O-P(HO)(O)-O-5'); 5'-triphosphate ((HO)2(O)P-O-(HO)(O)P-O-P(HO)(O)-O-5'); 5'-guanosine cap (7-methylated or non-methylated) (7m-G-O-5'-(HO)(O)P-O-(HO)(O)P-O-P(HO)(O)-O-5'); 5'-adenosine cap (App), and any modified or unmodified nucleotide cap structure (N-O-5'-(HO)(O)P-O-(HO)(O)P-O-P(HO)(O)-O-5'); 5'-monothiophosphate (phosphorothioate; (HO)2(S)P-O-5'); 5'-monodithiophosphate (phosphorodithioate; (HO)(HS)(S)P-O-5'); 5'-phosphorothiolate ((HO)2(O)P-S-5'); any additional combination of oxygen/sulfur replaced monophosphate, diphosphate and triphosphates (e.g. 5'-alpha-thiotriphosphate, 5'-gamma-thiotriphosphate, etc.), 5'-phosphoramidates ((HO)2(O)P-NH-5', (HO)(NH2)(O)P-O-5'), 5'-alkylphosphonates (R=alkyl=methyl, ethyl, isopropyl, propyl, etc., e.g. RP(OH)(O)-O-5', (OH)2(O)P-5'-CH2-), 5'-alkyletherphosphonates

Attorney's Docket No.: 14174-072W01

(R=alkylether=methoxymethyl (MeOCH₂-), ethoxymethyl, etc., e.g. RP(OH)(O)-O-5'-). (These modifications can also be used with the antisense strand of a double stranded iRNA.)

A single strand iRNA agent should be sufficiently long that it can enter the RISC and participate in RISC mediated cleavage of a target mRNA. A single strand iRNA agent is at least
14, and more preferably at least 15, 20, 25, 29, 35, 40, or 50 nucleotides in length. It is preferably
less than 200, 100, or 60 nucleotides in length.

Hairpin iRNA agents will have a duplex region equal to or at least 17, 18, 19, 29, 21, 22,
23, 24, or 25 nucleotide pairs. The duplex region will preferably be equal to or less than 200,
100, or 50, in length. Preferred ranges for the duplex region are 15-30, 17 to 23, 19 to 23, and 19
to 21 nucleotides pairs in length. The hairpin will preferably have a single strand overhang or
terminal unpaired region, preferably the 3', and preferably of the antisense side of the hairpin.
Preferred overhangs are 2-3 nucleotides in length.

A "double stranded (ds) iRNA agent" as used herein, is an iRNA agent which includes
more than one, and preferably two, strands in which interchain hybridization can form a region
of duplex structure.

The antisense strand of a double stranded iRNA agent should be equal to or at least, 14,
15, 16, 17, 18, 19, 25, 29, 40, or 60 nucleotides in length. It should be equal to or less than 200,
100, or 50, nucleotides in length. Preferred ranges are 17 to 25, 19 to 23, and 19 to 21
nucleotides in length.

The sense strand of a double stranded iRNA agent should be equal to or at least 14, 15,
16, 17, 18, 19, 25, 29, 40, or 60 nucleotides in length. It should be equal to or less than 200, 100,
or 50, nucleotides in length. Preferred ranges are 17 to 25, 19 to 23, and 19 to 21 nucleotides in
length.

The double strand portion of a double stranded iRNA agent should be equal to or at least,
14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 29, 40, or 60 nucleotide pairs in length. It should be
equal to or less than 200, 100, or 50, nucleotides pairs in length. Preferred ranges are 15-30, 17
to 23, 19 to 23, and 19 to 21 nucleotides pairs in length.

Attorney's Docket No.: 14174-072W01

In many embodiments, the ds iRNA agent is sufficiently large that it can be cleaved by an endogenous molecule, *e.g.*, by Dicer, to produce smaller ds iRNA agents, *e.g.*, sRNAs agents

It may be desirable to modify one or both of the antisense and sense strands of a double strand iRNA agent. In some cases they will have the same modification or the same class of modification but in other cases the sense and antisense strand will have different modifications, *e.g.*, in some cases it is desirable to modify only the sense strand. It may be desirable to modify only the sense strand, *e.g.*, to inactivate it, *e.g.*, the sense strand can be modified in order to inactivate the sense strand and prevent formation of an active sRNA/protein or RISC. This can be accomplished by a modification which prevents 5'-phosphorylation of the sense strand, *e.g.*, by modification with a 5'-O-methyl ribonucleotide (see Nykänen *et al.*, (2001) ATP requirements and small interfering RNA structure in the RNA interference pathway. Cell 107, 309-321.) Other modifications which prevent phosphorylation can also be used, *e.g.*, simply substituting the 5'-OH by H rather than O-Me. Alternatively, a large bulky group may be added to the 5'-phosphate turning it into a phosphodiester linkage, though this may be less desirable as phosphodiesterases can cleave such a linkage and release a functional sRNA 5'-end. Antisense strand modifications include 5' phosphorylation as well as any of the other 5' modifications discussed herein, particularly the 5' modifications discussed above in the section on single stranded iRNA molecules.

It is preferred that the sense and antisense strands be chosen such that the ds iRNA agent includes a single strand or unpaired region at one or both ends of the molecule. Thus, a ds iRNA agent contains sense and antisense strands, preferable paired to contain an overhang, *e.g.*, one or two 5' or 3' overhangs but preferably a 3' overhang of 2-3 nucleotides. Most embodiments will have a 3' overhang. Preferred sRNA agents will have single-stranded overhangs, preferably 3' overhangs, of 1 or preferably 2 or 3 nucleotides in length at each end. The overhangs can be the result of one strand being longer than the other, or the result of two strands of the same length being staggered. 5' ends are preferably phosphorylated.

Preferred lengths for the duplexed region is between 15 and 30, most preferably 18, 19, 20, 21, 22, and 23 nucleotides in length, *e.g.*, in the sRNA agent range discussed above. sRNA agents can resemble in length and structure the natural Dicer processed products from long

Attorney's Docket No.: 14174-072W01

dsRNAs. Embodiments in which the two strands of the sRNA agent are linked, *e.g.*, covalently linked are also included. Hairpin, or other single strand structures which provide the required double stranded region, and preferably a 3' overhang are also within the invention.

The isolated iRNA agents described herein, including ds iRNA agents and sRNA agents
5 can mediate silencing of a target RNA, *e.g.*, mRNA, *e.g.*, a transcript of a gene that encodes a protein. For convenience, such mRNA is also referred to herein as mRNA to be silenced. Such a gene is also referred to as a target gene. In general, the RNA to be silenced is an endogenous gene or a pathogen gene. In addition, RNAs other than mRNA, *e.g.*, tRNAs, and viral RNAs, can also be targeted.

10 As used herein, the phrase "mediates RNAi" refers to the ability to silence, in a sequence specific manner, a target RNA. While not wishing to be bound by theory, it is believed that silencing uses the RNAi machinery or process and a guide RNA, *e.g.*, an sRNA agent of 21 to 23 nucleotides.

As used herein, "specifically hybridizable" and "complementary" are terms which are
15 used to indicate a sufficient degree of complementarity such that stable and specific binding occurs between a compound of the invention and a target RNA molecule. Specific binding requires a sufficient degree of complementarity to avoid non-specific binding of the oligomeric compound to non-target sequences under conditions in which specific binding is desired, *i.e.*, under physiological conditions in the case of *in vivo* assays or therapeutic treatment, or in the
20 case of *in vitro* assays, under conditions in which the assays are performed. The non-target sequences typically differ by at least 5 nucleotides.

In one embodiment, an iRNA agent is "sufficiently complementary" to a target RNA, *e.g.*, a target mRNA, such that the iRNA agent silences production of protein encoded by the target mRNA. In another embodiment, the iRNA agent is "exactly complementary" (excluding
25 the RRMS containing subunit(s)) to a target RNA, *e.g.*, the target RNA and the iRNA agent anneal, preferably to form a hybrid made exclusively of Watson-Crick basepairs in the region of exact complementarity. A "sufficiently complementary" target RNA can include an internal region (*e.g.*, of at least 10 nucleotides) that is exactly complementary to a target RNA. Moreover, in some embodiments, the iRNA agent specifically discriminates a single-nucleotide

Attorney's Docket No.: 14174-072W01

difference. In this case, the iRNA agent only mediates RNAi if exact complementary is found in the region (*e.g.*, within 7 nucleotides of) the single-nucleotide difference.

As used herein, the term "oligonucleotide" refers to a nucleic acid molecule (RNA or DNA) preferably of length less than 100, 200, 300, or 400 nucleotides.

5 RNA agents discussed herein include otherwise unmodified RNA as well as RNA which have been modified, *e.g.*, to improve efficacy, and polymers of nucleoside surrogates. Unmodified RNA refers to a molecule in which the components of the nucleic acid, namely sugars, bases, and phosphate moieties, are the same or essentially the same as that which occur in nature, preferably as occur naturally in the human body. The art has referred to rare or unusual,
10 but naturally occurring, RNAs as modified RNAs, see, *e.g.*, Limbach *et al.*, (1994) Summary: the modified nucleosides of RNA, *Nucleic Acids Res.* 22: 2183-2196. Such rare or unusual RNAs, often termed modified RNAs (apparently because they are typically the result of a post transcriptionally modification) are within the term unmodified RNA, as used herein. Modified
15 RNA as used herein refers to a molecule in which one or more of the components of the nucleic acid, namely sugars, bases, and phosphate moieties, are different from that which occur in nature, preferably different from that which occurs in the human body. While they are referred to as modified "RNAs," they will of course, because of the modification, include molecules which are not RNAs. Nucleoside surrogates are molecules in which the ribophosphate backbone is replaced with a non-ribophosphate construct that allows the bases to be presented in the
20 correct spatial relationship such that hybridization is substantially similar to what is seen with a ribophosphate backbone, *e.g.*, non-charged mimics of the ribophosphate backbone. Examples of all of the above are discussed herein.

Much of the discussion below refers to single strand molecules. In many embodiments of the invention a double stranded iRNA agent, *e.g.*, a partially double stranded iRNA agent, is
25 required or preferred. Thus, it is understood that that double stranded structures (*e.g.* where two separate molecules are contacted to form the double stranded region or where the double stranded region is formed by intramolecular pairing (*e.g.*, a hairpin structure)) made of the single stranded structures described below are within the invention. Preferred lengths are described elsewhere herein.

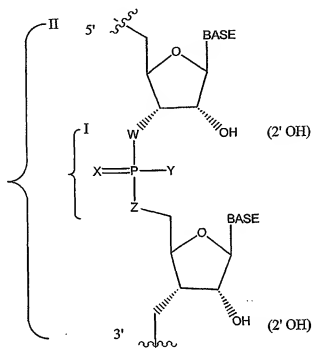
Attorney's Docket No.: 14174-072W01

As nucleic acids are polymers of subunits or monomers, many of the modifications described below occur at a position which is repeated within a nucleic acid, *e.g.*, a modification of a base, or a phosphate moiety, or the a non-linking O of a phosphate moiety. In some cases the modification will occur at all of the subject positions in the nucleic acid but in many, and
5 in fact in most cases it will not. By way of example, a modification may only occur at a 3' or 5' terminal position, may only occur in a terminal regions, *e.g.* at a position on a terminal nucleotide or in the last 2, 3, 4, 5, or 10 nucleotides of a strand. A modification may occur in a double strand region, a single strand region, or in both. A modification may occur only in the double strand region of an RNA or may only occur in a single strand region of an RNA. *E.g.*, a
10 phosphorothioate modification at a non-linking O position may only occur at one or both termini, may only occur in a terminal regions, *e.g.*, at a position on a terminal nucleotide or in the last 2, 3, 4, 5, or 10 nucleotides of a strand, or may occur in double strand and single strand regions, particularly at termini. The 5' end or ends can be phosphorylated.

In some embodiments it is particularly preferred, *e.g.*, to enhance stability, to include
15 particular bases in overhangs, or to include modified nucleotides or nucleotide surrogates, in single strand overhangs, *e.g.*, in a 5' or 3' overhang, or in both. *E.g.*, it can be desirable to include purine nucleotides in overhangs. In some embodiments all or some of the bases in a 3' or 5' overhang will be modified, *e.g.*, with a modification described herein. Modifications can include, *e.g.*, the use of modifications at the 2' OH group of the ribose sugar, *e.g.*, the use of
20 deoxyribonucleotides, *e.g.*, deoxythymidine, instead of ribonucleotides, and modifications in the phosphate group, *e.g.*, phosphothioate modifications. Overhangs need not be homologous with the target sequence.

Attorney's Docket No.: 14174-072W01

Modifications and nucleotide surrogates are discussed below.

FORMULA 1

5 The scaffold presented above in Formula 1 represents a portion of a ribonucleic acid. The basic components are the ribose sugar, the base, the terminal phosphates, and phosphate internucleotide linkers. Where the bases are naturally occurring bases, *e.g.*, adenine, uracil, guanine or cytosine, the sugars are the unmodified 2' hydroxyl ribose sugar (as depicted) and W, X, Y, and Z are all O. Formula 1 represents a naturally occurring unmodified

10 oligoribonucleotide.

Unmodified oligoribonucleotides may be less than optimal in some applications, *e.g.*, unmodified oligoribonucleotides can be prone to degradation by *e.g.*, cellular nucleases. Nucleases can hydrolyze nucleic acid phosphodiester bonds. However, chemical modifications

Attorney's Docket No.: 14174-072W01

to one or more of the above RNA components can confer improved properties, and, *e.g.*, can render oligoribonucleotides more stable to nucleases. Unmodified oligoribonucleotides may also be less than optimal in terms of offering tethering points for attaching ligands or other moieties to an iRNA agent.

5 Modified nucleic acids and nucleotide surrogates can include one or more of:

(i) alteration, *e.g.*, replacement, of one or both of the non-linking (X and Y) phosphate oxygens and/or of one or more of the linking (W and Z) phosphate oxygens (When the phosphate is in the terminal position, one of the positions W or Z will not link the phosphate to an additional element in a naturally occurring ribonucleic acid. However, for simplicity of
10 terminology, except where otherwise noted, the W position at the 5' end of a nucleic acid and the terminal Z position at the 3' end of a nucleic acid, are within the term "linking phosphate oxygens" as used herein.);

(ii) alteration, *e.g.*, replacement, of a constituent of the ribose sugar, *e.g.*, of the 2' hydroxyl on the ribose sugar, or wholesale replacement of the ribose sugar with a structure other
15 than ribose, *e.g.*, as described herein;

(iii) wholesale replacement of the phosphate moiety (bracket I) with "dephospho" linkers;

(iv) modification or replacement of a naturally occurring base;

(v) replacement or modification of the ribose-phosphate backbone (bracket II);

(vi) modification of the 3' end or 5' end of the RNA, *e.g.*, removal, modification or
20 replacement of a terminal phosphate group or conjugation of a moiety, *e.g.* a fluorescently labeled moiety, to either the 3' or 5' end of RNA.

The terms replacement, modification, alteration, and the like, as used in this context, do not imply any process limitation, *e.g.*, modification does not mean that one must start with a reference or naturally occurring ribonucleic acid and modify it to produce a modified ribonucleic
25 acid but rather modified simply indicates a difference from a naturally occurring molecule.

Attorney's Docket No.: 14174-072W01

It is understood that the actual electronic structure of some chemical entities cannot be adequately represented by only one canonical form (*i.e.* Lewis structure). While not wishing to be bound by theory, the actual structure can instead be some hybrid or weighted average of two or more canonical forms, known collectively as resonance forms or structures. Resonance structures are not discrete chemical entities and exist only on paper. They differ from one another only in the placement or "localization" of the bonding and nonbonding electrons for a particular chemical entity. It can be possible for one resonance structure to contribute to a greater extent to the hybrid than the others. Thus, the written and graphical descriptions of the embodiments of the present invention are made in terms of what the art recognizes as the predominant resonance form for a particular species. For example, any phosphoroamidate (replacement of a nonlinking oxygen with nitrogen) would be represented by $X = O$ and $Y = N$ in the above figure.

Specific modifications are discussed in more detail below.

The Phosphate Group

The phosphate group is a negatively charged species. The charge is distributed equally over the two non-linking oxygen atoms (*i.e.*, X and Y in Formula 1 above). However, the phosphate group can be modified by replacing one of the oxygens with a different substituent. One result of this modification to RNA phosphate backbones can be increased resistance of the oligoribonucleotide to nucleolytic breakdown. Thus while not wishing to be bound by theory, it can be desirable in some embodiments to introduce alterations which result in either an uncharged linker or a charged linker with unsymmetrical charge distribution.

Examples of modified phosphate groups include phosphorothioate, phosphoroselenates, borano phosphates, borano phosphate esters, hydrogen phosphonates, phosphoroamidates, alkyl or aryl phosphonates and phosphotriesters. Phosphorodithioates have both non-linking oxygens replaced by sulfur. Unlike the situation where only one of X or Y is altered, the phosphorus center in the phosphorodithioates is achiral which precludes the formation of oligoribonucleotides diastereomers. Diastereomer formation can result in a preparation in which the individual diastereomers exhibit varying resistance to nucleases. Further, the hybridization affinity of RNA containing chiral phosphate groups can be lower relative to the corresponding

Attorney's Docket No.: 14174-072W01

unmodified RNA species. Thus, while not wishing to be bound by theory, modifications to both X and Y which eliminate the chiral center, *e.g.* phosphorodithioate formation, may be desirable in that they cannot produce diastereomer mixtures. Thus, X can be any one of S, Se, B, C, H, N, or OR (R is alkyl or aryl). Thus Y can be any one of S, Se, B, C, H, N, or OR (R is alkyl or aryl). Replacement of X and/or Y with sulfur is preferred.

The phosphate linker can also be modified by replacement of a linking oxygen (*i.e.*, W or Z in Formula 1) with nitrogen (bridged phosphoramidates), sulfur (bridged phosphorothioates) and carbon (bridged methylenephosphonates). The replacement can occur at a terminal oxygen (position W (3') or position Z (5')). Replacement of W with carbon or Z with nitrogen is preferred.

Candidate agents can be evaluated for suitability as described below.

The Sugar Group

A modified RNA can include modification of all or some of the sugar groups of the ribonucleic acid. *E.g.*, the 2' hydroxyl group (OH) can be modified or replaced with a number of different "oxy" or "deoxy" substituents. While not being bound by theory, enhanced stability is expected since the hydroxyl can no longer be deprotonated to form a 2' alkoxide ion. The 2' alkoxide can catalyze degradation by intramolecular nucleophilic attack on the linker phosphorus atom. Again, while not wishing to be bound by theory, it can be desirable to some embodiments to introduce alterations in which alkoxide formation at the 2' position is not possible.

Examples of "oxy"-2' hydroxyl group modifications include alkoxy or aryloxy (OR, *e.g.*, R = H, alkyl, cycloalkyl, aryl, aralkyl, heteroaryl or sugar); polyethyleneglycols (PEG), $O(CH_2CH_2O)_nCH_2CH_2OR$; "locked" nucleic acids (LNA) in which the 2' hydroxyl is connected, *e.g.*, by a methylene bridge, to the 4' carbon of the same ribose sugar; O-AMINE (AMINE = NH_2 ; alkylamino, dialkylamino, heterocyclyl, arylamino, diaryl amino, heteroaryl amino, or diheteroaryl amino, ethylene diamine, polyamino) and aminoalkoxy, $O(CH_2)_nAMINE$, (*e.g.*, AMINE = NH_2 ; alkylamino, dialkylamino, heterocyclyl, arylamino, diaryl amino, heteroaryl amino, or diheteroaryl amino, ethylene diamine, polyamino). It is noteworthy that oligonucleotides containing only the methoxyethyl group (MOE), $(OCH_2CH_2OCH_3)$, a PEG

Attorney's Docket No.: 14174-072W01

derivative), exhibit nuclease stabilities comparable to those modified with the robust phosphorothioate modification.

“Deoxy” modifications include hydrogen (*i.e.* deoxyribose sugars, which are of particular relevance to the overhang portions of partially ds RNA); halo (*e.g.*, fluoro); amino (*e.g.* NH₂; 5 alkylamino, dialkylamino, heterocyclyl, arylamino, diaryl amino, heteroaryl amino, diheteroaryl amino, or amino acid); NH(CH₂CH₂NH)_nCH₂CH₂-AMINE (AMINE = NH₂; alkylamino, dialkylamino, heterocyclyl, arylamino, diaryl amino, heteroaryl amino, or diheteroaryl amino), -NHC(O)R (R = alkyl, cycloalkyl, aryl, aralkyl, heteroaryl or sugar), cyano; mercapto; alkyl-thio-alkyl; thioalkoxy; and alkyl, cycloalkyl, aryl, alkenyl and alkynyl, which may be optionally 10 substituted with *e.g.*, an amino functionality. Preferred substituents are 2'-methoxyethyl, 2'-OCH₃, 2'-O-allyl, 2'-C- allyl, and 2'-fluoro.

The sugar group can also contain one or more carbons that possess the opposite stereochemical configuration than that of the corresponding carbon in ribose. Thus, a modified RNA can include nucleotides containing *e.g.*, arabinose, as the sugar.

15 Modified RNAs can also include “abasic” sugars, which lack a nucleobase at C-1'. These abasic sugars can also be further contain modifications at one or more of the constituent sugar atoms.

To maximize nuclease resistance, the 2' modifications can be used in combination with one or more phosphate linker modifications (*e.g.*, phosphorothioate). The so-called “chimeric” 20 oligonucleotides are those that contain two or more different modifications.

The modification can also entail the wholesale replacement of a ribose structure with another entity at one or more sites in the iRNA agent. These modifications are described in section entitled Ribose Replacements for RRMSs.

Attorney's Docket No.: 14174-072W01

Candidate modifications can be evaluated as described below.

Replacement of the Phosphate Group

The phosphate group can be replaced by non-phosphorus containing connectors (*cf.* Bracket I in Formula 1 above). While not wishing to be bound by theory, it is believed that since
5 the charged phosphodiester group is the reaction center in nucleolytic degradation, its replacement with neutral structural mimics should impart enhanced nuclease stability. Again, while not wishing to be bound by theory, it can be desirable, in some embodiment, to introduce alterations in which the charged phosphate group is replaced by a neutral moiety.

Examples of moieties which can replace the phosphate group include siloxane, carbonate,
10 carboxymethyl, carbamate, amide, thioether, ethylene oxide linker, sulfonate, sulfonamide, thioformacetal, formacetal, oxime, methyleneimino, methylenemethylimino, methylenehydrazo, methylenedimethylhydrazo and methyleneoxymethylimino. Preferred replacements include the methylenecarbonylamino and methylenemethylimino groups.

Candidate modifications can be evaluated as described below.

15 Replacement of Ribophosphate Backbone

Oligonucleotide- mimicking scaffolds can also be constructed wherein the phosphate linker and ribose sugar are replaced by nuclease resistant nucleoside or nucleotide surrogates (see Bracket II of Formula 1 above). While not wishing to be bound by theory, it is believed that
20 the absence of a repetitively charged backbone diminishes binding to proteins that recognize polyanions (*e.g.* nucleases). Again, while not wishing to be bound by theory, it can be desirable in some embodiment, to introduce alterations in which the bases are tethered by a neutral surrogate backbone.

Examples include the morpholino, cyclobutyl, pyrrolidine and peptide nucleic acid (PNA) nucleoside surrogates. A preferred surrogate is a PNA surrogate.

25 Candidate modifications can be evaluated as described below.

Attorney's Docket No.: 14174-072W01

Terminal Modifications

The 3' and 5' ends of an oligonucleotide can be modified. Such modifications can be at the 3' end, 5' end or both ends of the molecule. They can include modification or replacement of an entire terminal phosphate or of one or more of the atoms of the phosphate group. *E.g.*, the 3' and 5' ends of an oligonucleotide can be conjugated to other functional molecular entities such as labeling moieties, *e.g.*, fluorophores (*e.g.*, pyrene, TAMRA, fluorescein, Cy3 or Cy5 dyes) or protecting groups (based *e.g.*, on sulfur, silicon, boron or ester). The functional molecular entities can be attached to the sugar through a phosphate group and/or a spacer. The terminal atom of the spacer can connect to or replace the linking atom of the phosphate group or the C-3' or C-5' O, N, S or C group of the sugar. Alternatively, the spacer can connect to or replace the terminal atom of a nucleotide surrogate (*e.g.*, PNAs). These spacers or linkers can include *e.g.*, $-(CH_2)_n-$, $-(CH_2)_nN-$, $-(CH_2)_nO-$, $-(CH_2)_nS-$, $O(CH_2CH_2O)_nCH_2CH_2OH$ (*e.g.*, $n = 3$ or 6), abasic sugars, amide, carboxy, amine, oxyamine, oximine, thioether, disulfide, thiourea, sulfonamide, or morpholino, or biotin and fluorescein reagents. When a spacer/phosphate-functional molecular entity-spacer/phosphate array is interposed between two strands of iRNA agents, this array can substitute for a hairpin RNA loop in a hairpin-type RNA agent. The 3' end can be an OH group. While not wishing to be bound by theory, it is believed that conjugation of certain moieties can improve transport, hybridization, and specificity properties. Again, while not wishing to be bound by theory, it may be desirable to introduce terminal alterations that improve nuclease resistance. Other examples of terminal modifications include dyes, intercalating agents (*e.g.* acridines), cross-linkers (*e.g.* psoralene, mitomycin C), porphyrins (TPPC4, texaphyrin, Sapphyrin), polycyclic aromatic hydrocarbons (*e.g.*, phenazine, dihydrophenazine), artificial endonucleases (*e.g.* EDTA), lipophilic carriers (*e.g.*, cholesterol, cholic acid, adamantane acetic acid, 1-pyrene butyric acid, dihydrotestosterone, 1,3-Bis-O(hexadecyl)glycerol, geranyloxyhexyl group, hexadecylglycerol, borneol, menthol, 1,3-propanediol, heptadecyl group, palmitic acid, myristic acid, O3-(oleoyl)lithocholic acid, O3-(oleoyl)cholenic acid, dimethoxytrityl, or phenoxazine) and peptide conjugates (*e.g.*, antennapedia peptide, Tat peptide), alkylating agents, phosphate, amino, mercapto, PEG (*e.g.*, PEG-40K), MPEG, [MPEG]₂, polyamino, alkyl, substituted alkyl, radiolabeled markers, enzymes, haptens (*e.g.* biotin), transport/absorption facilitators (*e.g.*, aspirin, vitamin E, folic acid), synthetic ribonucleases (*e.g.*, imidazole,

Attorney's Docket No.: 14174-072W01

bisimidazole, histamine, imidazole clusters, acridine-imidazole conjugates, Eu³⁺ complexes of tetraazamacrocycles).

Terminal modifications can be added for a number of reasons, including as discussed elsewhere herein to modulate activity or to modulate resistance to degradation. Terminal

5 modifications useful for modulating activity include modification of the 5' end with phosphate or phosphate analogs. *E.g.*, in preferred embodiments iRNA agents, especially antisense strands, are 5' phosphorylated or include a phosphoryl analog at the 5' prime terminus. 5'-phosphate

10 modifications include those which are compatible with RISC mediated gene silencing. Suitable modifications include: 5'-monophosphate ((HO)2(O)P-O-5'); 5'-diphosphate ((HO)2(O)P-O-P(O)(O)-O-5'); 5'-triphosphate ((HO)2(O)P-O-P(O)(O)-P(O)(O)-O-5'); 5'-guanosine cap (7-methylated or non-methylated) (7m-G-O-5'-(HO)(O)P-O-P(O)(O)-O-5'); 5'-adenosine cap (Appp), and any modified or unmodified nucleotide cap structure (N-O-5'-(HO)(O)P-O-P(O)(O)-O-5'); 5'-monothiophosphate (phosphorothioate; (HO)2(S)P-O-5'); 5'-monodithiophosphate (phosphorodithioate; (HO)(HS)(S)P-O-5'), 5'-phosphorothiolate ((HO)2(O)P-S-5'); any additional combination of oxygen/sulfur replaced

15 monophosphate, diphosphate and triphosphates (*e.g.* 5'-alpha-thiotriphosphate, 5'-gamma-thiotriphosphate, etc.), 5'-phosphoramidates ((HO)2(O)P-NH-5', (HO)(NH2)(O)P-O-5'), 5'-alkylphosphonates (R=alkyl=methyl, ethyl, isopropyl, propyl, etc., *e.g.* RP(OH)(O)-O-5', (OH)2(O)P-5'-CH2-), 5'-alkyletherphosphonates (R=alkylether=methoxymethyl (MeOCH2-),

20 ethoxymethyl, etc., *e.g.* RP(OH)(O)-O-5'-).

Terminal modifications can also be useful for monitoring distribution, and in such cases the preferred groups to be added include fluorophores, *e.g.*, fluorescein or an Alexa dye, *e.g.*, Alexa 488. Terminal modifications can also be useful for enhancing uptake, useful

25 modifications for this include cholesterol. Terminal modifications can also be useful for cross-linking an RNA agent to another moiety; modifications useful for this include mitomycin C.

Candidate modifications can be evaluated as described below.

Attorney's Docket No.: 14174-072W01

The Bases

Adenine, guanine, cytosine and uracil are the most common bases found in RNA. These bases can be modified or replaced to provide RNA's having improved properties. *E.g.*, nuclease resistant oligoribonucleotides can be prepared with these bases or with synthetic and natural nucleobases (*e.g.*, inosine, thymine, xanthine, hypoxanthine, nubarline, isoguanisine, or tubercidine) and any one of the above modifications. Alternatively, substituted or modified analogs of any of the above bases, *e.g.*, "unusual bases" and "universal bases," can be employed. Examples include without limitation 2-aminoadenine, 6-methyl and other alkyl derivatives of adenine and guanine, 2-propyl and other alkyl derivatives of adenine and guanine, 5-halouracil and cytosine, 5-propynyl uracil and cytosine, 6-azo uracil, cytosine and thymine, 5-uracil (pseudouracil), 4-thiouracil, 5-halouracil, 5-(2-aminopropyl)uracil, 5-amino allyl uracil, 8-halo, amino, thiol, thioalkyl, hydroxyl and other 8-substituted adenines and guanines, 5-trifluoromethyl and other 5-substituted uracils and cytosines, 7-methylguanine, 5-substituted pyrimidines, 6-azapyrimidines and N-2, N-6 and O-6 substituted purines, including 2-aminopropyladenine, 5-propynyluracil and 5-propynylcytosine, dihydrouracil, 3-deaza-5-azacytosine, 2-aminopurine, 5-alkyluracil, 7-alkylguanine, 5-alkyl cytosine, 7-deazaadenine, N6, N6-dimethyladenine, 2,6-diaminopurine, 5-amino-allyl-uracil, N3-methyluracil, substituted 1,2,4-triazoles, 2-pyridinone, 5-nitroindole, 3-nitropyrrole, 5-methoxyuracil, uracil-5-oxyacetic acid, 5-methoxycarbonylmethyluracil, 5-methyl-2-thiouracil, 5-methoxycarbonylmethyl-2-thiouracil, 5-methylaminomethyl-2-thiouracil, 3-(3-amino-3carboxypropyl)uracil, 3-methylcytosine, 5-methylcytosine, N⁴-acetyl cytosine, 2-thiocytosine, N6-methyladenine, N6-isopentyladenine, 2-methylthio-N6-isopentenyladenine, N-methylguanines, or O-alkylated bases. Further purines and pyrimidines include those disclosed in U.S. Pat. No. 3,687,808, those disclosed in the Concise Encyclopedia Of Polymer Science And Engineering, pages 858-859, Kroschwitz, J. I., ed. John Wiley & Sons, 1990, and those disclosed by Englisch *et al.*, *Angewandte Chemie, International Edition*, 1991, 30, 613.

Generally, base changes are less preferred for promoting stability, but they can be useful for other reasons, *e.g.*, some, *e.g.*, 2,6-diaminopurine and 2 amino purine, are fluorescent. Modified bases can reduce target specificity. This should be taken into consideration in the design of iRNA agents.

Candidate modifications can be evaluated as described below.

Evaluation of Candidate RNA's

One can evaluate a candidate RNA agent, *e.g.*, a modified RNA, for a selected property by exposing the agent or modified molecule and a control molecule to the appropriate conditions and evaluating for the presence of the selected property. For example, resistance to a degradant can be evaluated as follows. A candidate modified RNA (and preferably a control molecule, usually the unmodified form) can be exposed to degradative conditions, *e.g.*, exposed to a milieu, which includes a degradative agent, *e.g.*, a nuclease. *E.g.*, one can use a biological sample, *e.g.*, one that is similar to a milieu, which might be encountered, in therapeutic use, *e.g.*, blood or a cellular fraction, *e.g.*, a cell-free homogenate or disrupted cells. The candidate and control could then be evaluated for resistance to degradation by any of a number of approaches. For example, the candidate and control could be labeled, preferably prior to exposure, with, *e.g.*, a radioactive or enzymatic label, or a fluorescent label, such as Cy3 or Cy5. Control and modified RNA's can be incubated with the degradative agent, and optionally a control, *e.g.*, an inactivated, *e.g.*, heat inactivated, degradative agent. A physical parameter, *e.g.*, size, of the modified and control molecules are then determined. They can be determined by a physical method, *e.g.*, by polyacrylamide gel electrophoresis or a sizing column, to assess whether the molecule has maintained its original length, or assessed functionally. Alternatively, Northern blot analysis can be used to assay the length of an unlabeled modified molecule.

A functional assay can also be used to evaluate the candidate agent. A functional assay can be applied initially or after an earlier non-functional assay, (*e.g.*, assay for resistance to degradation) to determine if the modification alters the ability of the molecule to silence gene expression. For example, a cell, *e.g.*, a mammalian cell, such as a mouse or human cell, can be co-transfected with a plasmid expressing a fluorescent protein, *e.g.*, GFP, and a candidate RNA agent homologous to the transcript encoding the fluorescent protein (see, *e.g.*, WO 00/44914). For example, a modified dsRNA homologous to the GFP mRNA can be assayed for the ability to inhibit GFP expression by monitoring for a decrease in cell fluorescence, as compared to a control cell, in which the transfection did not include the candidate dsRNA, *e.g.*, controls with no agent added and/or controls with a non-modified RNA added. Efficacy of the candidate agent on

Attorney's Docket No.: 14174-072W01

gene expression can be assessed by comparing cell fluorescence in the presence of the modified and unmodified dsRNA agents.

In an alternative functional assay, a candidate dsRNA agent homologous to an endogenous mouse gene, preferably a maternally expressed gene, such as *c-mos*, can be injected into an immature mouse oocyte to assess the ability of the agent to inhibit gene expression *in vivo* (see, e.g., WO 01/36646). A phenotype of the oocyte, e.g., the ability to maintain arrest in metaphase II, can be monitored as an indicator that the agent is inhibiting expression. For example, cleavage of *c-mos* mRNA by a dsRNA agent would cause the oocyte to exit metaphase arrest and initiate parthenogenetic development (Colledge *et al.* Nature 370: 65-68, 1994; Hashimoto *et al.* Nature, 370:68-71, 1994). The effect of the modified agent on target RNA levels can be verified by Northern blot to assay for a decrease in the level of target mRNA, or by Western blot to assay for a decrease in the level of target protein, as compared to a negative control. Controls can include cells in which with no agent is added and/or cells in which a non-modified RNA is added.

References

General References

The oligoribonucleotides and oligoribonucleosides used in accordance with this invention may be with solid phase synthesis, see for example "Oligonucleotide synthesis, a practical approach", Ed. M. J. Gait, IRL Press, 1984; "Oligonucleotides and Analogues, A Practical Approach", Ed. F. Eckstein, IRL Press, 1991 (especially Chapter 1, Modern machine-aided methods of oligodeoxyribonucleotide synthesis, Chapter 2, Oligoribonucleotide synthesis, Chapter 3, 2'-O--Methyloligoribonucleotide- s: synthesis and applications, Chapter 4, Phosphorothioate oligonucleotides, Chapter 5, Synthesis of oligonucleotide phosphorodithioates, Chapter 6, Synthesis of oligo-2'-deoxyribonucleoside methylphosphonates, and. Chapter 7, Oligodeoxynucleotides containing modified bases. Other particularly useful synthetic procedures, reagents, blocking groups and reaction conditions are described in Martin, P., *Helv. Chim. Acta*, 1995, 78, 486-504; Beaucage, S. L. and Iyer, R. P., *Tetrahedron*, 1992, 48, 2223-2311 and Beaucage, S. L. and Iyer, R. P., *Tetrahedron*, 1993, 49, 6123-6194, or references referred to therein.

Attorney's Docket No.: 14174-072W01

Modification described in WO 00/44895, WO01/75164, or WO02/44321 can be used herein.

The disclosure of all publications, patents, and published patent applications listed herein are hereby incorporated by reference.

5 Phosphate Group References

The preparation of phosphinate oligoribonucleotides is described in U.S. Pat. No. 5,508,270. The preparation of alkyl phosphonate oligoribonucleotides is described in U.S. Pat. No. 4,469,863. The preparation of phosphoramidite oligoribonucleotides is described in U.S. Pat. No. 5,256,775 or U.S. Pat. No. 5,366,878. The preparation of phosphotriester oligoribonucleotides is described in U.S. Pat. No. 5,023,243. The preparation of borano phosphate oligoribonucleotide is described in U.S. Pat. Nos. 5,130,302 and 5,177,198. The preparation of 3'-Deoxy-3'-amino phosphoramidate oligoribonucleotides is described in U.S. Pat. No. 5,476,925. 3'-Deoxy-3'-methylenephosphonate oligoribonucleotides is described in An, H, *et al. J. Org. Chem.* **2001**, *66*, 2789-2801. Preparation of sulfur bridged nucleotides is described in Sproat *et al. Nucleosides Nucleotides* **1988**, *7*, 651 and Crosstick *et al. Tetrahedron Lett.* **1989**, *30*, 4693.

15 Sugar Group References

Modifications to the 2' modifications can be found in Verma, S. *et al. Annu. Rev. Biochem.* **1998**, *67*, 99-134 and all references therein. Specific modifications to the ribose can be found in the following references: 2'-fluoro (Kawasaki *et al.*, *J. Med. Chem.*, **1993**, *36*, 831-841), 2'-MOE (Martin, P. *Helv. Chim. Acta* **1996**, *79*, 1930-1938), "LNA" (Wengel, J. *Acc. Chem. Res.* **1999**, *32*, 301-310).

25 Replacement of the Phosphate Group References

Methylenemethylimino linked oligoribonucleosides, also identified herein as MMI linked oligoribonucleosides, methylenedimethylhydrazo linked oligoribonucleosides, also identified herein as MDH linked oligoribonucleosides, and methylenecarbonylamino linked oligonucleosides, also identified herein as amide-3 linked oligoribonucleosides, and

Attorney's Docket No.: 14174-072W01

methyleaminocarbonyl linked oligonucleosides, also identified herein as amide-4 linked oligoribonucleosides as well as mixed backbone compounds having, as for instance, alternating MMI and PO or PS linkages can be prepared as is described in U.S. Pat. Nos. 5,378,825, 5,386,023, 5,489,677 and in published PCT applications PCT/US92/04294 and
5 PCT/US92/04305 (published as WO 92/20822 WO and 92/20823, respectively). Fornacetal and thioformacetal linked oligoribonucleosides can be prepared as is described in U.S. Pat. Nos. 5,264,562 and 5,264,564. Bthylene oxide linked oligoribonucleosides can be prepared as is described in U.S. Pat. No. 5,223,618. Siloxane replacements are described in Cormier, J.F. *et al. Nucleic Acids Res.* 1988, 16, 4583. Carbonate replacements are described in Tittensor, J.R. *J. Chem. Soc. C* 1971, 1933. Carboxymethyl replacements are described in Edge, M.D. *et al. J. Chem. Soc. Perkin Trans. I* 1972, 1991. Carbamate replacements are described in Stirchak, E.P. *Nucleic Acids Res.* 1989, 17, 6129.

Replacement of the Phosphate-Ribose Backbone References

Cyclobutyl sugar surrogate compounds can be prepared as is described in U.S. Pat. No. 5,359,044. Pyrrolidine sugar surrogate can be prepared as is described in U.S. Pat. No. 5,519,134. Morpholino sugar surrogates can be prepared as is described in U.S. Pat. Nos. 5,142,047 and 5,235,033, and other related patent disclosures. Peptide Nucleic Acids (PNAs) are known per se and can be prepared in accordance with any of the various procedures referred to in Peptide Nucleic Acids (PNA): Synthesis, Properties and Potential Applications, Bioorganic &
20 Medicinal Chemistry, 1996, 4, 5-23. They may also be prepared in accordance with U.S. Pat. No. 5,539,083.

Terminal Modification References

Terminal modifications are described in Manoharan, M. *et al. Antisense and Nucleic Acid Drug Development* 12, 103-128 (2002) and references therein.

Bases References

N-2 substituted purine nucleoside amidites can be prepared as is described in U.S. Pat. No. 5,459,255. 3-Deaza purine nucleoside amidites can be prepared as is described in U.S. Pat. No. 5,457,191. 5,6-Substituted pyrimidine nucleoside amidites can be prepared as is described in

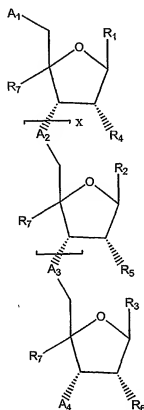
Attorney's Docket No.: 14174-072W01

U.S. Pat. No. 5,614,617. 5-Propynyl pyrimidine nucleoside amidites can be prepared as is described in U.S. Pat. No. 5,484,908. Additional references can be disclosed in the above section on base modifications.

Attorney's Docket No.: 14174-072W01

Preferred iRNA Agents

Preferred RNA agents have the following structure (see Formula 2 below):

FORMULA 2

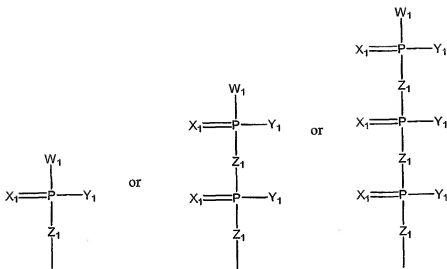
- 5 Referring to Formula 2 above, R^1 , R^2 , and R^3 are each, independently, H, (*i.e.* abasic nucleotides), adenine, guanine, cytosine and uracil, inosine, thymine, xanthine, hypoxanthine, nubularine, tubercidine, isoguanisine, 2-aminoadenine, 6-methyl and other alkyl derivatives of adenine and guanine, 2-propyl and other alkyl derivatives of adenine and guanine, 5-halouracil and cytosine, 5-propynyl uracil and cytosine, 6-azo uracil, cytosine and thymine, 5-uracil
- 10 (pseudouracil), 4-thiouracil, 5-halouracil, 5-(2-aminopropyl)uracil, 5-amino allyl uracil, 8-halo, amino, thiol, thioalkyl, hydroxyl and other 8-substituted adenines and guanines, 5-

Attorney's Docket No.: 14174-072W01

trifluoromethyl and other 5-substituted uracils and cytosines, 7-methylguanine, 5-substituted pyrimidines, 6-azapyrimidines and N-2, N-6 and O-6 substituted purines, including 2-aminopropyladenine, 5-propynyluracil and 5-propynylcytosine, dihydrouracil, 3-deaza-5-azacytosine, 2-aminopurine, 5-alkyluracil, 7-alkylguanine, 5-alkyl cytosine, 7-deazaadenine, 7-deazaguanine, N6, N6-dimethyladenine, 2,6-diaminopurine, 5-amino-allyl-uracil, N3-methyluracil, substituted 1,2,4-triazoles, 2-pyridinone, 5-nitroindole, 3-nitropyrrole, 5-methoxyuracil, uracil-5-oxyacetic acid, 5-methoxycarbonylmethyluracil, 5-methyl-2-thiouracil, 5-methoxycarbonylmethyl-2-thiouracil, 5-methylaminomethyl-2-thiouracil, 3-(3-amino-3carboxypropyl)uracil, 3-methylcytosine, 5-methylcytosine, N⁴-acetyl cytosine, 2-thiocytosine, N6-methyladenine, N6-isopentyladenine, 2-methylthio-N6-isopentenyladenine, N-methylguanines, or O-alkylated bases.

R⁴, R⁵, and R⁶ are each, independently, OR⁸, O(CH₂CH₂O)_mCH₂CH₂OR⁸; O(CH₂)_nR⁹; O(CH₂)_nOR⁹; H; halo; NH₂; NHR⁸; N(R⁸)₂; NH(CH₂CH₂NH)_mCH₂CH₂NHR⁹; NHC(O)R⁸; ; cyano; mercapto, SR⁸; alkyl-thio-alkyl; alkyl, aralkyl, cycloalkyl, aryl, heteroaryl, alkenyl, alkynyl, each of which may be optionally substituted with halo, hydroxy, oxo, nitro, haloalkyl, alkyl, alkaryl, aryl, aralkyl, alkoxy, aryloxy, amino, alkylamino, dialkylamino, heterocyclyl, arylamino, diaryl amino, heteroaryl amino, diheteroaryl amino, acylamino, alkylcarbonyl, arylcarbonyl, aminoalkyl, alkoxycarbonyl, carboxy, hydroxyalkyl, alkanesulfonyl, alkanesulfonamido, arenesulfonamido, aralkylsulfonamido, alkylcarbonyl, acyloxy, cyano, or ureido; or R⁴, R⁵, or R⁶ together combine with R⁷ to form an [-O-CH₂-] covalently bound bridge between the sugar 2' and 4' carbons.

Attorney's Docket No.: 14174-072W01

A¹ is:

5 ; H; OH; OCH₃; W¹; an abasic nucleotide; or absent;

(a preferred A¹, especially with regard to anti-sense strands, is chosen from 5'-monophosphate ((HO)₂(O)P-O-5'), 5'-diphosphate ((HO)₂(O)P-O-P(HO)(O)-O-5'), 5'-triphosphate ((HO)₂(O)P-O-(HO)(O)P-O-P(HO)(O)-O-5'), 5'-guanosine cap (7-methylated or non-methylated) (7m-G-O-5'-(HO)(O)P-O-(HO)(O)P-O-P(HO)(O)-O-5'), 5'-adenosine cap (A_{ppp}), and any modified or unmodified nucleotide cap structure (N-O-5'-(HO)(O)P-O-(HO)(O)P-O-P(HO)(O)-O-5'), 5'-monothiophosphate (phosphorothioate; (HO)₂(S)P-O-5'), 5'-monodithiophosphate (phosphorodithioate; (HO)(HS)(S)P-O-5'), 5'-phosphorothiolate ((HO)₂(O)P-S-5'); any additional combination of oxygen/sulfur replaced monophosphate, diphosphate and triphosphates (e.g. 5'-alpha-thiotriphosphate, 5'-gamma-thiotriphosphate, etc.), 5'-phosphoramidates ((HO)₂(O)P-NH-5', (HO)(NH₂)(O)P-O-5'), 5'-alkylphosphonates (R=alkyl=methyl, ethyl, isopropyl, propyl, etc., e.g. RP(OH)(O)-O-5', (OH)₂(O)P-5'-CH₂-), 5'-

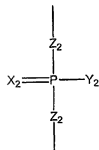
10

15

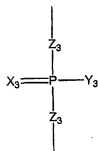
Attorney's Docket No.: 14174-072W01

alkyletherphosphonates (R=alkylether=methoxymethyl (MeOCH₂-), ethoxymethyl, etc., e.g. RP(OH)(O)-O-S'-)).

A² is:

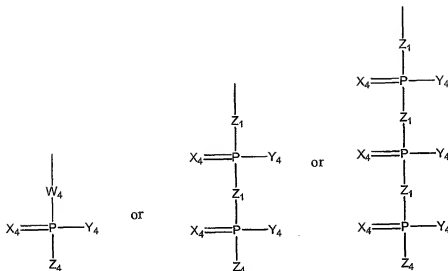


A³ is:



; and

Attorney's Docket No.: 14174-072W01

 A^4 is:

5 ; H; Z^4 ; an inverted nucleotide; an abasic nucleotide; or absent.

W^1 is OH , $(CH_2)_n R^{10}$, $(CH_2)_n NHR^{10}$, $(CH_2)_n OR^{10}$, $(CH_2)_n SR^{10}$, $O(CH_2)_n R^{10}$, $O(CH_2)_n OR^{10}$, $O(CH_2)_n NR^{10}$, $O(CH_2)_n SR^{10}$, $O(CH_2)_n SS(CH_2)_n OR^{10}$, $O(CH_2)_n C(O)OR^{10}$, $NH(CH_2)_n R^{10}$, $NH(CH_2)_n NR^{10}$, $NH(CH_2)_n OR^{10}$, $NH(CH_2)_n SR^{10}$, $S(CH_2)_n R^{10}$, $S(CH_2)_n NR^{10}$, $S(CH_2)_n OR^{10}$, $S(CH_2)_n SR^{10}$, $O(CH_2 CH_2 O)_m CH_2 CH_2 OR^{10}$, $O(CH_2 CH_2 O)_m CH_2 CH_2 NHR^{10}$, $10 NH(CH_2 CH_2 NH)_m CH_2 CH_2 NHR^{10}$, $Q-R^{10}$, $O-Q-R^{10}$, $N-Q-R^{10}$, $S-Q-R^{10}$ or $-O-$. W^4 is O , CH_2 , NH , or S .

X^1 , X^2 , X^3 , and X^4 are each, independently, O or S .

Y^1 , Y^2 , Y^3 , and Y^4 are each, independently, OH , O^- , OR^8 , S , Se , BH_3^- , H , NHR^9 , $N(R^9)_2$, alkyl, cycloalkyl, aralkyl, aryl, or heteroaryl, each of which may be optionally substituted.

15 Z^1 , Z^2 , and Z^3 are each independently O , CH_2 , NH , or S . Z^4 is OH , $(CH_2)_n R^{10}$, $(CH_2)_n NHR^{10}$, $(CH_2)_n OR^{10}$, $(CH_2)_n SR^{10}$, $O(CH_2)_n R^{10}$, $O(CH_2)_n OR^{10}$, $O(CH_2)_n NR^{10}$,

Attorney's Docket No.: 14174-072W01

$O(CH_2)_nSR^{10}$, $O(CH_2)_nSS(CH_2)_nOR^{10}$, $O(CH_2)_nC(O)OR^{10}$, $NH(CH_2)_nR^{10}$, $NH(CH_2)_nNR^{10}$
 $;NH(CH_2)_nOR^{10}$, $NH(CH_2)_nSR^{10}$, $S(CH_2)_nR^{10}$, $S(CH_2)_nNR^{10}$, $S(CH_2)_nOR^{10}$, $S(CH_2)_nSR^{10}$
 $O(CH_2CH_2O)_mCH_2CH_2OR^{10}$, $O(CH_2CH_2O)_mCH_2CH_2NHR^{10}$,
 $NH(CH_2CH_2NH)_mCH_2CH_2NHR^{10}$, $Q-R^{10}$, $O-Q-R^{10}$, $N-Q-R^{10}$, $S-Q-R^{10}$.

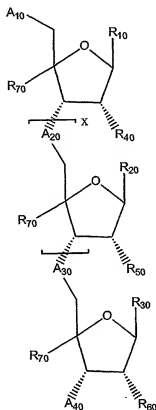
5 x is 5-100, chosen to comply with a length for an RNA agent described herein.

R^7 is H; or is together combined with R^4 , R^5 , or R^6 to form an $[-O-CH_2-]$ covalently bound bridge between the sugar $2'$ and $4'$ carbons.

R^8 is alkyl, cycloalkyl, aryl, aralkyl, heterocyclyl, heteroaryl, amino acid, or sugar; R^9 is NH_2 , alkylamino, dialkylamino, heterocyclyl, arylamino, diaryl amino, heteroaryl amino,
 10 diheteroaryl amino, or amino acid; and R^{10} is H; fluorophore (pyrene, TAMRA, fluorescein, Cy3 or Cy5 dyes); sulfur, silicon, boron or ester protecting group; intercalating agents (e.g. acridines), cross-linkers (e.g. psoralene, mitomycin C), porphyrins (TPPC4, texaphyrin, Sapphyrin), polycyclic aromatic hydrocarbons (e.g., phenazine, dihydrophenazine), artificial endonucleases (e.g. EDTA), lipophilic carriers (cholesterol, cholic acid, adamantane acetic acid, 1-pyrene butyric
 15 acid, dihydrotestosterone, 1,3-Bis-O(hexadecyl)glycerol, geranyloxyhexyl group, hexadecylglycerol, borneol, menthol, 1,3-propanediol, heptadecyl group, palmitic acid, myristic acid, O3-(oleoyl)lithocholic acid, O3-(oleoyl)cholenic acid, dimethoxytrityl, or phenoxazine) and peptide conjugates (e.g., antennapedia peptide, Tat peptide), alkylating agents, phosphate, amino, mercapto, PEG (e.g., PEG-40K), MPEG, [MPEG]₂, polyamino; alkyl, cycloalkyl, aryl, aralkyl,
 20 heteroaryl; radiolabelled markers, enzymes, haptens (e.g. biotin), transport/absorption facilitators (e.g., aspirin, vitamin E, folic acid), synthetic ribonucleases (e.g., imidazole, bisimidazole, histamine, imidazole clusters, acridine-imidazole conjugates, Eu^{3+} complexes of tetraazamacrocycles); or an RNA agent. m is 0-1,000,000, and n is 0-20. Q is a spacer selected from the group consisting of abasic sugar, amide, carboxy, oxyamine, oximine, thioether,
 25 disulfide, thiourea, sulfonamide, or morpholino, biotin or fluorescein reagents.

Attorney's Docket No.: 14174-072/W01

Preferred RNA agents in which the entire phosphate group has been replaced have the following structure (see Formula 3 below):

FORMULA 3

- Referring to Formula 3, A¹⁰-A⁴⁰ is L-G-L; A¹⁰ and/or A⁴⁰ may be absent, in which L is a linker, wherein one or both L may be present or absent and is selected from the group consisting of CH₂(CH₂)₆; N(CH₂)₆; O(CH₂)₆; S(CH₂)₆. G is a functional group selected from the group consisting of siloxane, carbonate, carboxymethyl, carbamate, amide, thioether, ethylene oxide linker, sulfonate, sulfonamide, thioformacetal, formacetal, oxime, methyleneimino, methylenemethylimino, methylenehydrazo, methylenedimethylhydrazo and methyleneoxymethylimino.

Attorney's Docket No.: 14174-072W01

R^{10} , R^{20} , and R^{30} are each, independently, H, (*i.e.* abasic nucleotides), adenine, guanine, cytosine and uracil, inosine, thymine, xanthine, hypoxanthine, nubularine, tubercidine, isoguanisine, 2-aminoadenine, 6-methyl and other alkyl derivatives of adenine and guanine, 2-propyl and other alkyl derivatives of adenine and guanine, 5-halouracil and cytosine, 5-propynyl
 5 uracil and cytosine, 6-azo uracil, cytosine and thymine, 5-uracil (pseudouracil), 4-thiouracil, 5-halouracil, 5-(2-aminopropyl)uracil, 5-amino allyl uracil, 8-halo, amino, thiol, thioalkyl, hydroxyl and other 8-substituted adenines and guanines, 5-trifluoromethyl and other 5-substituted uracils and cytosines, 7-methylguanine, 5-substituted pyrimidines, 6-azapyrimidines and N-2, N-6 and O-6 substituted purines, including 2-aminopropyladenine, 5-propynyluracil
 10 and 5-propynylcytosine, dihydrouracil, 3-deaza-5-azacytosine, 2-aminopurine, 5-alkyluracil, 7-alkylguanine, 5-alkyl cytosine, 7-deazaadenine, 7-deazaguanine, N6, N6-dimethyladenine, 2,6-diaminopurine, 5-amino-allyl-uracil, N3-methyluracil substituted 1,2,4-triazoles, 2-pyridinone, 5-nitroindole, 3-nitropyrrrole, 5-methoxyuracil, uracil-5-oxyacetic acid, 5-methoxycarbonylmethyluracil, 5-methyl-2-thiouracil, 5-methoxycarbonylmethyl-2-thiouracil, 5-methylaminomethyl-2-thiouracil, 3-(3-amino-3-carboxypropyl)uracil, 3-methylcytosine, 5-methylcytosine, N^4 -acetyl cytosine, 2-thiocytosine, N6-methyladenine, N6-isopentyladenine, 2-methylthio-N6-isopentyladenine, N-methylguanines, or O-alkylated bases.

R^{40} , R^{50} , and R^{60} are each, independently, OR^8 , $O(CH_2CH_2O)_mCH_2CH_2OR^8$, $O(CH_2)_nR^9$, $O(CH_2)_nOR^9$, H; halo; NH_2 ; NHR^8 ; $N(R^8)_2$; $NH(CH_2CH_2NH)_mCH_2CH_2R^9$; $NHC(O)R^8$; cyano; mercapto, SR^7 ; alkyl-thio-alkyl; alkyl, aralkyl, cycloalkyl, aryl, heteroaryl, alkenyl, alkynyl, each
 20 of which may be optionally substituted with halo, hydroxy, oxo, nitro, haloalkyl, alkyl, alkaryl, aryl, aralkyl, alkoxy, aryloxy, amino, alkylamino, dialkylamino, heterocyclyl, arylamino, diaryl amino, heteroaryl amino, diheteroaryl amino, acylamino, alkylcarbonyl, arylcarbonyl, aminoalkyl, alkoxy carbonyl, carboxy, hydroxyalkyl, alkanesulfonyl, alkanesulfonamido, arenesulfonamido, aralkylsulfonamido, alkylcarbonyl, acyloxy, cyano, and ureido groups; or R^{40} , R^{50} , or R^{60} together combine with R^{70} to form an $[-O-CH_2-]$ covalently bound bridge between the
 25 sugar 2' and 4' carbons.

x is 5-100 or chosen to comply with a length for an RNA agent described herein.

Attorney's Docket No.: 14174-072W01

R^{70} is H; or is together combined with R^{40} , R^{50} , or R^{60} to form an $[-O-CH_2-]$ covalently bound bridge between the sugar 2' and 4' carbons.

R^8 is alkyl, cycloalkyl, aryl, aralkyl, heterocyclyl, heteroaryl, amino acid, or sugar; and R^9 is NH_2 , alkylamino, dialkylamino, heterocyclyl, arylamino, diaryl amino, heteroaryl amino, diheteroaryl amino, or amino acid. m is 0-1,000,000, n is 0-20, and g is 0-2.

Preferred nucleoside surrogates have the following structure (see Formula 4 below):



FORMULA 4

10 S is a nucleoside surrogate selected from the group consisting of morpholino, cyclobutyl, pyrrolidine and peptide nucleic acid. L is a linker and is selected from the group consisting of $CH_2(CH_2)_6$, $N(CH_2)_6$, $O(CH_2)_6$, $S(CH_2)_6$, $-C(O)(CH_2)_n$ - or may be absent. M is an amide bond; sulfonamide; sulfinate; phosphate group; modified phosphate group as described herein; or may be absent.

15 R^{100} , R^{200} , and R^{300} are each, independently, H (*i.e.*, abasic nucleotides), adenine, guanine, cytosine and uracil, inosine, thymine, xanthine, hypoxanthine, nubularine, tubercidine, isoguanisine, 2-aminoadenine, 6-methyl and other alkyl derivatives of adenine and guanine, 2-propyl and other alkyl derivatives of adenine and guanine, 5-halouracil and cytosine, 5-propynyl uracil and cytosine, 6-azo uracil, cytosine and thymine, 5-uracil (pseudouracil), 4-thiouracil, 5-halouracil, 5-(2-aminopropyl)uracil, 5-amino allyl uracil, 8-halo, amino, thiol, thioalkyl, hydroxyl and other 8-substituted adenines and guanines, 5-trifluoromethyl and other 5-substituted uracils and cytosines, 7-methylguanine, 5-substituted pyrimidines, 6-azapyrimidines and N-2, N-6 and O-6 substituted purines, including 2-aminopropyladenine, 5-propynyluracil and 5-propynylcytosine, dihydrouracil, 3-deaza-5-azacytosine, 2-aminopurine, 5-alkyluracil, 7-alkylguanine, 5-alkyl cytosine, 7-deazaadenine, 7-deazaguanine, N6, N6-dimethyladenine, 2,6-diaminopurine, 5-amino-allyl-uracil, N3-methyluracil substituted 1, 2, 4, triazoles, 2-pyridinones, 5-nitroindole, 3-nitropyrrole, 5-methoxyuracil, uracil-5-oxyacetic acid, 5-

Attorney's Docket No.: 14174-072W01

methoxycarbonylmethyluracil, 5-methyl-2-thiouracil, 5-methoxycarbonylmethyl-2-thiouracil, 5-methylaminomethyl-2-thiouracil, 3-(3-amino-3-carboxypropyl)uracil, 3-methylcytosine, 5-methylcytosine, N⁴-acetyl cytosine, 2-thiocytosine, N⁶-methyladenine, N⁶-isopentyladenine, 2-methylthio-N⁶-isopentyladenine, N-methylguanines, or O-alkylated bases.

- 5 x is 5-100, or chosen to comply with a length for an RNA agent described herein; and g is 0-2.

Nuclease resistant monomers

- 10 An RNA, e.g., an iRNA agent, can incorporate a nuclease resistant monomer (NRM), such as those described herein and those described in copending, co-owned United States Provisional Application Serial No. 60/469,612, filed on May 9, 2003, and International Application No. PCT/US04/07070, both of which are hereby incorporated by reference.

- 15 In addition, the invention includes iRNA agents having an NRM and another element described herein. E.g., the invention includes an iRNA agent described herein, e.g., a palindromic iRNA agent, an iRNA agent having a non canonical pairing, an iRNA agent which targets a gene described herein, e.g., a gene active in the liver, an iRNA agent having an architecture or structure described herein, an iRNA associated with an amphipathic delivery agent described herein, an iRNA associated with a drug delivery module described herein, an iRNA agent administered as described herein, or an iRNA agent formulated as described herein, which also incorporates an NRM.

- 20 An iRNA agent can include monomers which have been modified so as to inhibit degradation, e.g., by nucleases, e.g., endonucleases or exonucleases, found in the body of a subject. These monomers are referred to herein as NRMs, or nuclease resistance promoting monomers or modifications. In many cases these modifications will modulate other properties of the iRNA agent as well, e.g., the ability to interact with a protein, e.g., a transport protein, e.g., serum albumin, or a member of the RISC (RNA-induced Silencing Complex), or the ability of 25 the first and second sequences to form a duplex with one another or to form a duplex with another sequence, e.g., a target molecule.

Attorney's Docket No.: 14174-072W01

While not wishing to be bound by theory, it is believed that modifications of the sugar, base, and/or phosphate backbone in an iRNA agent can enhance endonuclease and exonuclease resistance, and can enhance interactions with transporter proteins and one or more of the functional components of the RISC complex. Preferred modifications are those that increase exonuclease and endonuclease resistance and thus prolong the half-life of the iRNA agent prior to interaction with the RISC complex, but at the same time do not render the iRNA agent resistant to endonuclease activity in the RISC complex. Again, while not wishing to be bound by any theory, it is believed that placement of the modifications at or near the 3' and/or 5' end of antisense strands can result in iRNA agents that meet the preferred nuclease resistance criteria delineated above. Again, still while not wishing to be bound by any theory, it is believed that placement of the modifications at e.g., the middle of a sense strand can result in iRNA agents that are relatively less likely to undergo off-targeting.

Modifications described herein can be incorporated into any double-stranded RNA and RNA-like molecule described herein, e.g., an iRNA agent. An iRNA agent may include a duplex comprising a hybridized sense and antisense strand, in which the antisense strand and/or the sense strand may include one or more of the modifications described herein. The anti sense strand may include modifications at the 3' end and/or the 5' end and/or at one or more positions that occur 1-6 (e.g., 1-5, 1-4, 1-3, 1-2) nucleotides from either end of the strand. The sense strand may include modifications at the 3' end and/or the 5' end and/or at any one of the intervening positions between the two ends of the strand. The iRNA agent may also include a duplex comprising two hybridized antisense strands. The first and/or the second antisense strand may include one or more of the modifications described herein. Thus, one and/or both antisense strands may include modifications at the 3' end and/or the 5' end and/or at one or more positions that occur 1-6 (e.g., 1-5, 1-4, 1-3, 1-2) nucleotides from either end of the strand. Particular configurations are discussed below.

Modifications that can be useful for producing iRNA agents that meet the preferred nuclease resistance criteria delineated above can include one or more of the following chemical and/or stereochemical modifications of the sugar, base, and/or phosphate backbone:

Attorney's Docket No.: 14174-072W01

(i) chiral (S_P) thioates. Thus, preferred NRMs include nucleotide dimers with an enriched or pure for a particular chiral form of a modified phosphate group containing a heteroatom at the nonbridging position, e.g., S_P or R_P , at the position X , where this is the position normally occupied by the oxygen. The atom at X can also be S, Se, Nr_2 , or Br_3 . When X is S, enriched or
5 chirally pure S_P linkage is preferred. Enriched means at least 70, 80, 90, 95, or 99% of the preferred form. Such NRMs are discussed in more detail below;

(ii) attachment of one or more cationic groups to the sugar, base, and/or the phosphorus atom of a phosphate or modified phosphatic backbone moiety. Thus, preferred NRMs include monomers at the terminal position derivatized at a cationic group. As the 5' end of an antisense
10 sequence should have a terminal -OH or phosphate group this NRM is preferably not used at the 5' end of an anti-sense sequence. The group should be attached at a position on the base which minimizes interference with H bond formation and hybridization, e.g., away from the face which interacts with the complementary base on the other strand, e.g., at the 5' position of a pyrimidine or a 7-position of a purine. These are discussed in more detail below;

(iii) nonphosphate linkages at the termini. Thus, preferred NRMs include Non-phosphate linkages, e.g., a linkage of 4 atoms which confers greater resistance to cleavage than does a phosphate bond. Examples include 3' $CH_2-NCH_3-O-CH_2-5'$ and 3' $CH_2-NH-(O=)-CH_2-5'$;

(iv) 3'-bridging thiophosphates and 5'-bridging thiophosphates. Thus, preferred NRM's can included these structures;

(v) L-RNA, 2'-5' linkages, inverted linkages, a-nucleosides. Thus, other preferred
20 NRM's include: L nucleosides and dimeric nucleotides derived from L-nucleosides; 2'-5' phosphate, non-phosphate and modified phosphate linkages (e.g., thiophosphates, phosphoramidates and boronophosphates); dimers having inverted linkages, e.g., 3'-3' or 5'-5' linkages; monomers having an alpha linkage at the 1' site on the sugar, e.g., the structures
25 described herein having an alpha linkage;

(vi) conjugate groups. Thus, preferred NRM's can include e.g., a targeting moiety or a conjugated ligand described herein conjugated with the monomer, e.g., through the sugar, base, or backbone;

Attorney's Docket No.: 14174-072W01

(vi) abasic linkages. Thus, preferred NRM's can include an abasic monomer, e.g., an abasic monomer as described herein (e.g., a nucleobaseless monomer); an aromatic or heterocyclic or polyheterocyclic aromatic monomer as described herein.; and

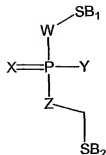
(vii) 5'-phosphonates and 5'-phosphate prodrugs. Thus, preferred NRM's include
5 monomers, preferably at the terminal position, e.g., the 5' position, in which one or more atoms of the phosphate group is derivatized with a protecting group, which protecting group or groups, are removed as a result of the action of a component in the subject's body, e.g., a carboxyesterase or an enzyme present in the subject's body. E.g., a phosphate prodrug in which a carboxy
10 esterase cleaves the protected molecule resulting in the production of a thioate anion which attacks a carbon adjacent to the O of a phosphate and resulting in the production of an unprotected phosphate.

One or more different NRM modifications can be introduced into an iRNA agent or into a sequence of an iRNA agent. An NRM modification can be used more than once in a sequence or in an iRNA agent. As some NRM's interfere with hybridization the total number
15 incorporated, should be such that acceptable levels of iRNA agent duplex formation are maintained.

In some embodiments NRM modifications are introduced into the terminal the cleavage site or in the cleavage region of a sequence (a sense strand or sequence) which does not target a desired sequence or gene in the subject. This can reduce off-target silencing.

20 *Chiral S_P Thioates*

A modification can include the alteration, e.g., replacement, of one or both of the non-linking (X and Y) phosphate oxygens and/or of one or more of the linking (W and Z) phosphate oxygens. Formula X below depicts a phosphate moiety linking two sugar/sugar surrogate-base moieties, SB₁ and SB₂.

FORMULA X

In certain embodiments, one of the non-linking phosphate oxygens in the phosphate backbone moiety (X and Y) can be replaced by any one of the following: S, Se, BR₃ (R is hydrogen, alkyl, aryl, etc.), C (i.e., an alkyl group, an aryl group, etc.), H, NR₂ (R is hydrogen, alkyl, aryl, etc.), or OR (R is alkyl or aryl). The phosphorus atom in an unmodified phosphate group is achiral. However, replacement of one of the non-linking oxygens with one of the above atoms or groups of atoms renders the phosphorus atom chiral; in other words a phosphorus atom in a phosphate group modified in this way is a stereogenic center. The stereogenic phosphorus atom can possess either the "R" configuration (herein R_P) or the "S" configuration (herein S_P). Thus if 60% of a population of stereogenic phosphorus atoms have the R_P configuration, then the remaining 40% of the population of stereogenic phosphorus atoms have the S_P configuration.

In some embodiments, iRNA agents, having phosphate groups in which a phosphate non-linking oxygen has been replaced by another atom or group of atoms, may contain a population of stereogenic phosphorus atoms in which at least about 50% of these atoms (e.g., at least about 60% of these atoms, at least about 70% of these atoms, at least about 80% of these atoms, at least about 90% of these atoms, at least about 95% of these atoms, at least about 98% of these atoms, at least about 99% of these atoms) have the S_P configuration. Alternatively, iRNA agents having phosphate groups in which a phosphate non-linking oxygen has been replaced by another atom or group of atoms may contain a population of stereogenic phosphorus atoms in which at least about 50% of these atoms (e.g., at least about 60% of these atoms, at least about 70% of these

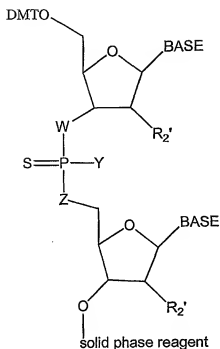
Attorney's Docket No.: 14174-072W01

atoms, at least about 80% of these atoms, at least about 90% of these atoms, at least about 95% of these atoms, at least about 98% of these atoms, at least about 99% of these atoms) have the R_P configuration. In other embodiments, the population of stereogenic phosphorus atoms may have the S_P configuration and may be substantially free of stereogenic phosphorus atoms having the R_P configuration. In still other embodiments, the population of stereogenic phosphorus atoms may have the R_P configuration and may be substantially free of stereogenic phosphorus atoms having the S_P configuration. As used herein, the phrase "substantially free of stereogenic phosphorus atoms having the R_P configuration" means that moieties containing stereogenic phosphorus atoms having the R_P configuration cannot be detected by conventional methods known in the art (chiral HPLC, 1H NMR analysis using chiral shift reagents, etc.). As used herein, the phrase "substantially free of stereogenic phosphorus atoms having the S_P configuration" means that moieties containing stereogenic phosphorus atoms having the S_P configuration cannot be detected by conventional methods known in the art (chiral HPLC, 1H NMR analysis using chiral shift reagents, etc.).

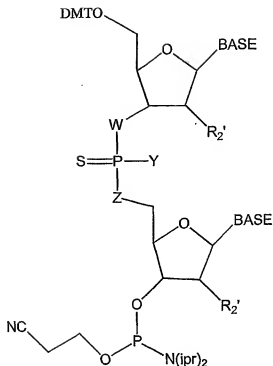
In a preferred embodiment, modified iRNA agents contain a phosphorothioate group, i.e., a phosphate groups in which a phosphate non-linking oxygen has been replaced by a sulfur atom. In an especially preferred embodiment, the population of phosphorothioate stereogenic phosphorus atoms may have the S_P configuration and be substantially free of stereogenic phosphorus atoms having the R_P configuration.

Phosphorothioates may be incorporated into iRNA agents using dimers e.g., formulas X-1 and X-2. The former can be used to introduce phosphorothioate

Attorney's Docket No.: 14174-072W01



X-1



X-2

at the 3' end of a strand, while the latter can be used to introduce this modification at the 5' end or at a position that occurs e.g., 1, 2, 3, 4, 5, or 6 nucleotides from either end of the strand. In the
 5 above formulas, Y can be 2-cyanoethoxy, W and Z can be O, R₂' can be, e.g., a substituent that can impart the C-3 endo configuration to the sugar (e.g., OH, F, OCH₃), DMT is dimethoxytrityl, and "BASE" can be a natural, unusual, or a universal base.

X-1 and X-2 can be prepared using chiral reagents or directing groups that can result in phosphorothioate-containing dimers having a population of stereogenic phosphorus atoms
 10 having essentially only the R_P configuration (i.e., being substantially free of the S_P configuration) or only the S_P configuration (i.e., being substantially free of the R_P configuration). Alternatively, dimers can be prepared having a population of stereogenic phosphorus atoms in which about

Attorney's Docket No.: 14174-072W01

50% of the atoms have the R_P configuration and about 50% of the atoms have the S_P configuration. Dimers having stereogenic phosphorus atoms with the R_P configuration can be identified and separated from dimers having stereogenic phosphorus atoms with the S_P configuration using e.g., enzymatic degradation and/or conventional chromatography techniques.

Cationic Groups

Modifications can also include attachment of one or more cationic groups to the sugar, base, and/or the phosphorus atom of a phosphate or modified phosphate backbone moiety. A cationic group can be attached to any atom capable of substitution on a natural, unusual or universal base. A preferred position is one that does not interfere with hybridization, i.e., does not interfere with the hydrogen bonding interactions needed for base pairing. A cationic group can be attached e.g., through the C2' position of a sugar or analogous position in a cyclic or acyclic sugar surrogate. Cationic groups can include e.g., protonated amino groups, derived from e.g., O-AMINE (AMINE = NH_2 ; alkylamino, dialkylamino, heterocyclyl, arylamino, diaryl amino, heteroaryl amino, or diheteroaryl amino, ethylene diamine, polyamino); aminoalkoxy, e.g., $O(CH_2)_n$ AMINE, (e.g., AMINE = NH_2 ; alkylamino, dialkylamino, heterocyclyl, arylamino, diaryl amino, heteroaryl amino, or diheteroaryl amino, ethylene diamine, polyamino); amino (e.g. NH_2 ; alkylamino, dialkylamino, heterocyclyl, arylamino, diaryl amino, heteroaryl amino, diheteroaryl amino, or amino acid); or $NH(CH_2CH_2NH)_nCH_2CH_2$ -AMINE (AMINE = NH_2 ; alkylamino, dialkylamino, heterocyclyl, arylamino, diaryl amino, heteroaryl amino, or diheteroaryl amino).

Nonphosphate Linkages

Modifications can also include the incorporation of nonphosphate linkages at the 5' and/or 3' end of a strand. Examples of nonphosphate linkages which can replace the phosphate group include methyl phosphonate, hydroxylamino, siloxane, carbonate, carboxymethyl, carbamate, amide, thioether, ethylene oxide linker, sulfonate, sulfonamide, thioformacetal, formacetal, oxime, methyleneimino, methylenemethylimino, methylenehydrazo, methylenedimethylhydrazo and methyleneoxymethylimino. Preferred replacements include the methyl phosphonate and hydroxylamino groups.

Attorney's Docket No.: 14174-072W01

3'-bridging thiophosphates and 5'-bridging thiophosphates; locked-RNA, 2'-5' linkages, inverted linkages, α -nucleosides; conjugate groups; abasic linkages; and 5'-phosphonates and 5'-phosphate prodrugs

Referring to formula X above, modifications can include replacement of one of the
5 bridging or linking phosphate oxygens in the phosphate backbone moiety (W and Z). Unlike the situation where only one of X or Y is altered, the phosphorus center in the phosphorodithioates is achiral which precludes the formation of iRNA agents containing a stereogenic phosphorus atom.

Modifications can also include linking two sugars via a phosphate or modified phosphate
10 group through the 2' position of a first sugar and the 5' position of a second sugar. Also contemplated are inverted linkages in which both a first and second sugar are each linked through the respective 3' positions. Modified RNA's can also include "abasic" sugars, which lack a nucleobase at C-1'. The sugar group can also contain one or more carbons that possess the opposite stereochemical configuration than that of the corresponding carbon in ribose. Thus, a
15 modified iRNA agent can include nucleotides containing e.g., arabinose, as the sugar. In another subset of this modification, the natural, unusual, or universal base may have the α -configuration. Modifications can also include L-RNA.

Modifications can also include 5'-phosphonates, e.g., $P(O)(O)_2-X-C^{5'}\text{-sugar}$ ($X = CH_2$, CF_2 , CHF and 5'-phosphate prodrugs, e.g., $P(O)[OCH_2CH_2SC(O)R]_2CH_2C^{5'}\text{-sugar}$. In the
20 latter case, the prodrug groups may be decomposed via reaction first with carboxy esterases. The remaining ethyl thiolate group via intramolecular S_N2 displacement can depart as episulfide to afford the underivatized phosphate group.

Modification can also include the addition of conjugating groups described elsewhere
herein, which are preferably attached to an iRNA agent through any amino group available for
25 conjugation.

Nuclease resistant modifications include some which can be placed only at the terminus and others which can go at any position. Generally the modifications that can inhibit hybridization so it is preferably to use them only in terminal regions, and preferable to not use

Attorney's Docket No.: 14174-072W01

them at the cleavage site or in the cleavage region of a sequence which targets a subject sequence or gene.. The can be used anywhere in a sense sequence, provided that sufficient hybridization between the two sequences of the iRNA agent is maintained. In some embodiments it is desirable to put the NRM at the cleavage site or in the cleavage region of a sequence which does not target a subject sequence or gene, as it can minimize off-target silencing.

In addition, an iRNA agent described herein can have an overhang which does not form a duplex structure with the other sequence of the iRNA agent—it is an overhang, but it does hybridize, either with itself, or with another nucleic acid, other than the other sequence of the iRNA agent.

In most cases, the nuclease-resistance promoting modifications will be distributed differently depending on whether the sequence will target a sequence in the subject (often referred to as an anti-sense sequence) or will not target a sequence in the subject (often referred to as a sense sequence). If a sequence is to target a sequence in the subject, modifications which interfere with or inhibit endonuclease cleavage should not be inserted in the region which is subject to RISC mediated cleavage, e.g., the cleavage site or the cleavage region (As described in Elbashir *et al.*, 2001, *Genes and Dev.* 15: 188, hereby incorporated by reference, cleavage of the target occurs about in the middle of a 20 or 21 nt guide RNA, or about 10 or 11 nucleotides upstream of the first nucleotide which is complementary to the guide sequence. As used herein cleavage site refers to the nucleotide on either side of the cleavage site, on the target or on the iRNA agent strand which hybridizes to it. Cleavage region means an nucleotide with 1, 2, or 3 nucleotides of the cleave site, in either direction.)

Such modifications can be introduced into the terminal regions, e.g., at the terminal position or with 2, 3, 4, or 5 positions of the terminus, of a sequence which targets or a sequence which does not target a sequence in the subject.

An iRNA agent can have a first and a second strand chosen from the following:

a first strand which does not target a sequence and which has an NRM modification at or within 1, 2, 3, 4, 5, or 6 positions from the 3' end;

Attorney's Docket No.: 14174-072W01

a first strand which does not target a sequence and which has an NRM modification at or within 1, 2, 3, 4, 5, or 6 positions from the 5' end;

a first strand which does not target a sequence and which has an NRM modification at or within 1, 2, 3, 4, 5, or 6 positions from the 3' end and which has a NRM modification at or
5 within 1, 2, 3, 4, 5, or 6 positions from the 5' end;

a first strand which does not target a sequence and which has an NRM modification at the cleavage site or in the cleavage region;

a first strand which does not target a sequence and which has an NRM modification at the cleavage site or in the cleavage region and one or more of an NRM modification at or within 1,
10 2, 3, 4, 5, or 6 positions from the 3' end, a NRM modification at or within 1, 2, 3, 4, 5, or 6 positions from the 5' end, or NRM modifications at or within 1, 2, 3, 4, 5, or 6 positions from both the 3' and the 5' end; and

a second strand which targets a sequence and which has an NRM modification at or within 1, 2, 3, 4, 5, or 6 positions from the 3' end;

a second strand which targets a sequence and which has an NRM modification at or within 1, 2, 3, 4, 5, or 6 positions from the 5' end (5' end NRM modifications are preferentially
15 not at the terminus but rather at a position 1, 2, 3, 4, 5, or 6 away from the 5' terminus of an antisense strand);

a second strand which targets a sequence and which has an NRM modification at or within 1, 2, 3, 4, 5, or 6 positions from the 3' end and which has a NRM modification at or
20 within 1, 2, 3, 4, 5, or 6 positions from the 5' end;

a second strand which targets a sequence and which preferably does not have an NRM modification at the cleavage site or in the cleavage region;

a second strand which targets a sequence and which does not have an NRM modification at the cleavage site or in the cleavage region and one or more of an NRM modification at or
25 within 1, 2, 3, 4, 5, or 6 positions from the 3' end, a NRM modification at or within 1, 2, 3, 4, 5, or 6 positions from the 5' end, or NRM modifications at or within 1, 2, 3, 4, 5, or 6 positions

Attorney's Docket No.: 14174-072W01

from both the 3' and the 5' end (5' end NRM modifications are preferentially not at the terminus but rather at a position 1, 2, 3, 4, 5, or 6 away from the 5' terminus of an antisense strand).

An iRNA agent can also target two sequences and can have a first and second strand chosen from:

5 a first strand which targets a sequence and which has an NRM modification at or within 1, 2, 3, 4, 5, or 6 positions from the 3' end;

 a first strand which targets a sequence and which has an NRM modification at or within 1, 2, 3, 4, 5, or 6 positions from the 5' end (5' end NRM modifications are preferentially not at the terminus but rather at a position 1, 2, 3, 4, 5, or 6 away from the 5' terminus of an antisense
10 strand);

 a first strand which targets a sequence and which has an NRM modification at or within 1, 2, 3, 4, 5, or 6 positions from the 3' end and which has a NRM modification at or within 1, 2, 3, 4, 5, or 6 positions from the 5' end;

 a first strand which targets a sequence and which preferably does not have an NRM
15 modification at the cleavage site or in the cleavage region;

 a first strand which targets a sequence and which does not have an NRM modification at the cleavage site or in the cleavage region and one or more of an NRM modification at or within 1, 2, 3, 4, 5, or 6 positions from the 3' end, a NRM modification at or within 1, 2, 3, 4, 5, or 6
20 positions from the 5' end, or NRM modifications at or within 1, 2, 3, 4, 5, or 6 positions from both the 3' and the 5' end (5' end NRM modifications are preferentially not at the terminus but rather at a position 1, 2, 3, 4, 5, or 6 away from the 5' terminus of an antisense strand) and

 a second strand which targets a sequence and which has an NRM modification at or within 1, 2, 3, 4, 5, or 6 positions from the 3' end;

 a second strand which targets a sequence and which has an NRM modification at or
25 within 1, 2, 3, 4, 5, or 6 positions from the 5' end (5' end NRM modifications are preferentially not at the terminus but rather at a position 1, 2, 3, 4, 5, or 6 away from the 5' terminus of an antisense strand);

Attorney's Docket No.: 14174-072W01

a second strand which targets a sequence and which has an NRM modification at or within 1, 2, 3, 4, 5, or 6 positions from the 3' end and which has a NRM modification at or within 1, 2, 3, 4, 5, or 6 positions from the 5' end;

5 a second strand which targets a sequence and which preferably does not have an NRM modification at the cleavage site or in the cleavage region;

a second strand which targets a sequence and which does not have an NRM modification at the cleavage site or in the cleavage region and one or more of an NRM modification at or within 1, 2, 3, 4, 5, or 6 positions from the 3' end, a NRM modification at or within 1, 2, 3, 4, 5, or 6 positions from the 5' end, or NRM modifications at or within 1, 2, 3, 4, 5, or 6 positions from both the 3' and the 5' end (5' end NRM modifications are preferentially not at the terminus but rather at a position 1, 2, 3, 4, 5, or 6 away from the 5' terminus of an antisense strand).

10

Ribose Mimics

An RNA, e.g., an iRNA agent, can incorporate a ribose mimic, such as those described herein and those described in copending co-owned United States Provisional Application Serial No. 60/454,962, filed on March 13, 2003, and International Application No. PCT/US04/07070, both of which are hereby incorporated by reference.

15

In addition, the invention includes iRNA agents having a ribose mimic and another element described herein. E.g., the invention includes an iRNA agent described herein, e.g., a palindromic iRNA agent, an iRNA agent having a non canonical pairing, an iRNA agent which targets a gene described herein, e.g., a gene active in the liver, an iRNA agent having an architecture or structure described herein, an iRNA associated with an amphipathic delivery agent described herein, an iRNA associated with a drug delivery module described herein, an iRNA agent administered as described herein, or an iRNA agent formulated as described herein, which also incorporates a ribose mimic.

20

Thus, an aspect of the invention features an iRNA agent that includes a secondary hydroxyl group, which can increase efficacy and/or confer nuclease resistance to the agent. Nucleases, e.g., cellular nucleases, can hydrolyze nucleic acid phosphodiester bonds, resulting in partial or complete degradation of the nucleic acid. The secondary hydroxy group confers

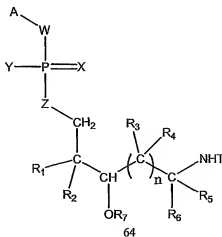
25

Attorney's Docket No.: 14174-072W01

nuclease resistance to an iRNA agent by rendering the iRNA agent less prone to nuclease degradation relative to an iRNA which lacks the modification. While not wishing to be bound by theory, it is believed that the presence of a secondary hydroxyl group on the iRNA agent can act as a structural mimic of a 3' ribose hydroxyl group, thereby causing it to be less susceptible to degradation.

The secondary hydroxyl group refers to an "OH" radical that is attached to a carbon atom substituted by two other carbons and a hydrogen. The secondary hydroxyl group that confers nuclease resistance as described above can be part of any acyclic carbon-containing group. The hydroxyl may also be part of any cyclic carbon-containing group, and preferably one or more of the following conditions is met (1) there is no ribose moiety between the hydroxyl group and the terminal phosphate group or (2) the hydroxyl group is not on a sugar moiety which is coupled to a base. The hydroxyl group is located at least two bonds (e.g., at least three bonds away, at least four bonds away, at least five bonds away, at least six bonds away, at least seven bonds away, at least eight bonds away, at least nine bonds away, at least ten bonds away, etc.) from the terminal phosphate group phosphorus of the iRNA agent. In preferred embodiments, there are five intervening bonds between the terminal phosphate group phosphorus and the secondary hydroxyl group.

Preferred iRNA agent delivery modules with five intervening bonds between the terminal phosphate group phosphorus and the secondary hydroxyl group have the following structure (see formula Y below):



(Y)

Referring to formula Y, A is an iRNA agent, including any iRNA agent described herein. The iRNA agent may be connected directly or indirectly (e.g., through a spacer or linker) to "W" of the phosphate group. These spacers or linkers can include e.g., $-(CH_2)_n-$, $-(CH_2)_nN-$, $-(CH_2)_nO-$, $-(CH_2)_nS-$, $O(CH_2CH_2O)_nCH_2CH_2OH$ (e.g., $n = 3$ or 6), abasic sugars, amide, carboxy, amine, oxyamine, oximine, thioether, disulfide, thiourea, sulfonamide, or morpholino, or biotin and fluorescein reagents.

The iRNA agents can have a terminal phosphate group that is unmodified (e.g., W, X, Y, and Z are O) or modified. In a modified phosphate group, W and Z can be independently NH, O, or S; and X and Y can be independently S, Se, BH_3^- , C_1-C_6 alkyl, C_6-C_{10} aryl, H, O, O^- , alkoxy or amino (including alkylamino, arylamino, etc.). Preferably, W, X and Z are O and Y is S.

R_1 and R_3 are each, independently, hydrogen; or C_1-C_{100} alkyl, optionally substituted with hydroxyl, amino, halo, phosphate or sulfate and/or may be optionally inserted with N, O, S, alkenyl or alkynyl.

R_2 is hydrogen; C_1-C_{100} alkyl, optionally substituted with hydroxyl, amino, halo, phosphate or sulfate and/or may be optionally inserted with N, O, S, alkenyl or alkynyl; or, when n is 1, R_2 may be taken together with R_4 or R_6 to form a ring of 5-12 atoms.

R_4 is hydrogen; C_1-C_{100} alkyl, optionally substituted with hydroxyl, amino, halo, phosphate or sulfate and/or may be optionally inserted with N, O, S, alkenyl or alkynyl; or, when n is 1, R_4 may be taken together with R_2 or R_5 to form a ring of 5-12 atoms.

R_5 is hydrogen, C_1-C_{100} alkyl optionally substituted with hydroxyl, amino, halo, phosphate or sulfate and/or may be optionally inserted with N, O, S, alkenyl or alkynyl; or, when n is 1, R_5 may be taken together with R_4 to form a ring of 5-12 atoms.

R_6 is hydrogen, C_1-C_{100} alkyl, optionally substituted with hydroxyl, amino, halo, phosphate or sulfate and/or may be optionally inserted with N, O, S, alkenyl or alkynyl, or, when n is 1, R_6 may be taken together with R_2 to form a ring of 6-10 atoms;

Attorney's Docket No.: 14174-072W01

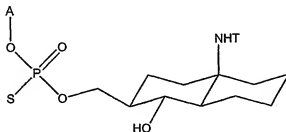
R_7 is hydrogen, C_1 - C_{100} alkyl, or $C(O)(CH_2)_qC(O)NHR_9$; T is hydrogen or a functional group; n and q are each independently 1-100; R_8 is C_1 - C_{10} alkyl or C_6 - C_{10} aryl; and R_9 is hydrogen, C_1 - C_{10} alkyl, C_6 - C_{10} aryl or a solid support agent.

Preferred embodiments may include one or more of the following subsets of iRNA agent delivery modules.

In one subset of RNAi agent delivery modules, A can be connected directly or indirectly through a terminal 3' or 5' ribose sugar carbon of the RNA agent.

In another subset of RNAi agent delivery modules, X, W, and Z are O and Y is S.

In still yet another subset of RNAi agent delivery modules, n is 1, and R_2 and R_6 are taken together to form a ring containing six atoms and R_4 and R_5 are taken together to form a ring containing six atoms. Preferably, the ring system is a *trans*-decalin. For example, the RNAi agent delivery module of this subset can include a compound of Formula (Y-1):



The functional group can be, for example, a targeting group (e.g., a steroid or a carbohydrate), a reporter group (e.g., a fluorophore), or a label (an isotopically labelled moiety). The targeting group can further include protein binding agents, endothelial cell targeting groups (e.g., RGD peptides and mimetics), cancer cell targeting groups (e.g., folate Vitamin B12, Biotin), bone cell targeting groups (e.g., bisphosphonates, polyglutamates, polyaspartates), multivalent mannose (for e.g., macrophage testing), lactose, galactose, N-acetyl-galactosamine, monoclonal antibodies, glycoproteins, lectins, melanotropin, or thyrotropin.

Attorney's Docket No.: 14174-072W01

As can be appreciated by the skilled artisan, methods of synthesizing the compounds of the formulae herein will be evident to those of ordinary skill in the art. The synthesized compounds can be separated from a reaction mixture and further purified by a method such as column chromatography, high pressure liquid chromatography, or recrystallization.

- 5 Additionally, the various synthetic steps may be performed in an alternate sequence or order to give the desired compounds. Synthetic chemistry transformations and protecting group methodologies (protection and deprotection) useful in synthesizing the compounds described herein are known in the art and include, for example, those such as described in R. Larock, *Comprehensive Organic Transformations*, VCH Publishers (1989); T.W. Greene and P.G.M. Wuts, *Protective Groups in Organic Synthesis*, 2d. Ed., John Wiley and Sons (1991); L. Fieser and M. Fieser, *Fieser and Fieser's Reagents for Organic Synthesis*, John Wiley and Sons (1994); and L. Paquette, ed., *Encyclopedia of Reagents for Organic Synthesis*, John Wiley and Sons (1995), and subsequent editions thereof.

Ribose Replacement Monomer Subunits

- 15 iRNA agents can be modified in a number of ways which can optimize one or more characteristics of the iRNA agent. An RNA agent, e.g., an iRNA agent can include a ribose replacement monomer subunit (RRMS), such as those described herein and those described in one or more of United States Provisional Application Serial No. 60/493,986, filed on August 8, 2003, which is hereby incorporated by reference; United States Provisional Application Serial No. 60/494,597, filed on August 11, 2003, which is hereby incorporated by reference; United States Provisional Application Serial No. 60/506,341, filed on September 26, 2003, which is hereby incorporated by reference; United States Provisional Application Serial No. 60/158,453, filed on November 7, 2003, which is hereby incorporated by reference; and International Application No. PCT/US04/07070, filed March 8, 2004, which is hereby
25 incorporated by reference.

In addition, the invention includes iRNA agents having a RRMS and another element described herein. E.g., the invention includes an iRNA agent described herein, e.g., a palindromic iRNA agent, an iRNA agent having a non canonical pairing, an iRNA agent which targets a gene described herein, e.g., a gene active in the liver, an iRNA agent having an

Attorney's Docket No.: 14174-072W01

architecture or structure described herein, an iRNA associated with an amphipathic delivery agent described herein, an iRNA associated with a drug delivery module described herein, an iRNA agent administered as described herein, or an iRNA agent formulated as described herein, which also incorporates a RRMS.

- 5 The ribose sugar of one or more ribonucleotide subunits of an iRNA agent can be replaced with another moiety, e.g., a non-carbohydrate (preferably cyclic) carrier. A ribonucleotide subunit in which the ribose sugar of the subunit has been so replaced is referred to herein as a ribose replacement modification subunit (RRMS). A cyclic carrier may be a carbocyclic ring system, i.e., all ring atoms are carbon atoms, or a heterocyclic ring system, i.e.,
- 10 one or more ring atoms may be a heteroatom, e.g., nitrogen, oxygen, sulfur. The cyclic carrier may be a monocyclic ring system, or may contain two or more rings, e.g. fused rings. The cyclic carrier may be a fully saturated ring system, or it may contain one or more double bonds.

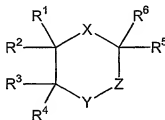
- The carriers further include (i) at least two "backbone attachment points" and (ii) at least one "tethering attachment point." A "backbone attachment point" as used herein refers to a
- 15 functional group, e.g. a hydroxyl group, or generally, a bond available for, and that is suitable for incorporation of the carrier into the backbone, e.g., the phosphate, or modified phosphate, e.g., sulfur containing, backbone, of a ribonucleic acid. A "tethering attachment point" as used herein
- refers to a constituent ring atom of the cyclic carrier, e.g., a carbon atom or a heteroatom (distinct from an atom which provides a backbone attachment point), that connects a selected moiety.
- 20 The moiety can be, e.g., a ligand, e.g., a targeting or delivery moiety, or a moiety which alters a physical property, e.g., lipophilicity, of an iRNA agent. Optionally, the selected moiety is connected by an intervening tether to the cyclic carrier. Thus, it will include a functional group, e.g., an amino group, or generally, provide a bond, that is suitable for incorporation or tethering of another chemical entity, e.g., a ligand to the constituent ring.

- 25 Incorporation of one or more RRMSs described herein into an RNA agent, e.g., an iRNA agent, particularly when tethered to an appropriate entity, can confer one or more new properties to the RNA agent and/or alter, enhance or modulate one or more existing properties in the RNA molecule. E.g., it can alter one or more of lipophilicity or nuclease resistance. Incorporation of one or more RRMSs described herein into an iRNA agent can, particularly when the RRMS is

Attorney's Docket No.: 14174-072W01

tethered to an appropriate entity, modulate, e.g., increase, binding affinity of an iRNA agent to a target mRNA, change the geometry of the duplex form of the iRNA agent, alter distribution or target the iRNA agent to a particular part of the body, or modify the interaction with nucleic acid binding proteins (e.g., during RISC formation and strand separation).

- 5 Accordingly, in one aspect, the invention features, an iRNA agent preferably comprising a first strand and a second strand, wherein at least one subunit having a formula (R-1) is incorporated into at least one of said strands.



(R-1)

- 10 Referring to formula (R-1), X is $\text{N}(\text{CO})\text{R}^7$, NR^7 or CH_2 ; Y is NR^8 , O, S, CR^9R^{10} , or absent; and Z is $\text{CR}^{11}\text{R}^{12}$ or absent.

- Each of R^1 , R^2 , R^3 , R^4 , R^9 , and R^{10} is, independently, H, OR^a , OR^b , $(\text{CH}_2)_n\text{OR}^a$, or $(\text{CH}_2)_n\text{OR}^b$, provided that at least one of R^1 , R^2 , R^3 , R^4 , R^9 , and R^{10} is OR^a or OR^b and that at least one of R^1 , R^2 , R^3 , R^4 , R^9 , and R^{10} is $(\text{CH}_2)_n\text{OR}^a$, or $(\text{CH}_2)_n\text{OR}^b$ (when the RRMS is terminal, one of R^1 , R^2 , R^3 , R^4 , R^9 , and R^{10} will include R^a and one will include R^b ; when the RRMS is internal, two of R^1 , R^2 , R^3 , R^4 , R^9 , and R^{10} will each include an R^b); further provided that preferably OR^a may only be present with $(\text{CH}_2)_n\text{OR}^b$ and $(\text{CH}_2)_n\text{OR}^a$ may only be present with OR^b .

- Each of R^5 , R^6 , R^{11} , and R^{12} is, independently, H, $\text{C}_1\text{-C}_6$ alkyl optionally substituted with 1-3 R^{13} , or $\text{C}(\text{O})\text{NHR}^7$; or R^5 and R^{11} together are $\text{C}_3\text{-C}_8$ cycloalkyl optionally substituted with R^{14} .

Attorney's Docket No.: 14174-072W01

R^7 is C_1 - C_{20} alkyl substituted with NR^5R^d ; R^8 is C_1 - C_6 alkyl; R^{13} is hydroxy, C_1 - C_4 alkoxy, or halo; and R^{14} is NR^5R^7 .

5

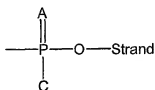
R^a is:



; and

10

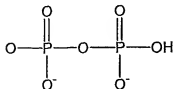
R^b is:



Each of A and C is, independently, O or S.

B is OH, O', or

Attorney's Docket No.: 14174-072W01



R^c is H or C1-C6 alkyl; R^d is H or a ligand; and n is 1-4.

In a preferred embodiment the ribose is replaced with a pyrroline scaffold, and X is
 5 $\text{N}(\text{CO})\text{R}^7$ or NR^7 , Y is CR^9R^{10} , and Z is absent.

In other preferred embodiments the ribose is replaced with a piperidine scaffold, and X is
 $\text{N}(\text{CO})\text{R}^7$ or NR^7 , Y is CR^9R^{10} , and Z is $\text{CR}^{11}\text{R}^{12}$.

In other preferred embodiments the ribose is replaced with a piperazine scaffold, and X is
 $\text{N}(\text{CO})\text{R}^7$ or NR^7 , Y is NR^8 , and Z is $\text{CR}^{11}\text{R}^{12}$.

10 In other preferred embodiments the ribose is replaced with a morpholino scaffold, and X
 is $\text{N}(\text{CO})\text{R}^7$ or NR^7 , Y is O, and Z is $\text{CR}^{11}\text{R}^{12}$.

In other preferred embodiments the ribose is replaced with a decalin scaffold, and X
 is CH_2 ; Y is CR^9R^{10} ; and Z is $\text{CR}^{11}\text{R}^{12}$; and R^5 and R^{11} together are C^6 cycloalkyl.

In other preferred embodiments the ribose is replaced with a decalin/indane scaffold and ,
 15 and X is CH_2 ; Y is CR^9R^{10} ; and Z is $\text{CR}^{11}\text{R}^{12}$; and R^5 and R^{11} together are C^5 cycloalkyl.

In other preferred embodiments, the ribose is replaced with a hydroxyproline scaffold.

RRMSs described herein may be incorporated into any double-stranded RNA-like
 molecule described herein, e.g., an iRNA agent. An iRNA agent may include a duplex
 comprising a hybridized sense and antisense strand, in which the antisense strand and/or the
 20 sense strand may include one or more of the RRMSs described herein. An RRMS can be
 introduced at one or more points in one or both strands of a double-stranded iRNA agent. An

Attorney's Docket No.: 14174-072W01

RRMS can be placed at or near (within 1, 2, or 3 positions) of the 3' or 5' end of the sense strand or at near (within 2 or 3 positions of) the 3' end of the antisense strand. In some embodiments it is preferred to not have an RRMS at or near (within 1, 2, or 3 positions of) the 5' end of the antisense strand. An RRMS can be internal, and will preferably be positioned in regions not
5 critical for antisense binding to the target.

In an embodiment, an iRNA agent may have an RRMS at (or within 1, 2, or 3 positions of) the 3' end of the antisense strand. In an embodiment, an iRNA agent may have an RRMS at (or within 1, 2, or 3 positions of) the 3' end of the antisense strand and at (or within 1, 2, or 3 positions of) the 3' end of the sense strand. In an embodiment, an iRNA agent may have an
10 RRMS at (or within 1, 2, or 3 positions of) the 3' end of the antisense strand and an RRMS at the 5' end of the sense strand, in which both ligands are located at the same end of the iRNA agent.

In certain embodiments, two ligands are tethered, preferably, one on each strand and are hydrophobic moieties. While not wishing to be bound by theory, it is believed that pairing of the hydrophobic ligands can stabilize the iRNA agent *via* intermolecular van der Waals interactions.

15 In an embodiment, an iRNA agent may have an RRMS at (or within 1, 2, or 3 positions of) the 3' end of the antisense strand and an RRMS at the 5' end of the sense strand, in which both RRMSs may share the same ligand (e.g., cholic acid) via connection of their individual tethers to separate positions on the ligand. A ligand shared between two proximal RRMSs is referred to herein as a "hairpin ligand."

20 In other embodiments, an iRNA agent may have an RRMS at the 3' end of the sense strand and an RRMS at an internal position of the sense strand. An iRNA agent may have an RRMS at an internal position of the sense strand; or may have an RRMS at an internal position of the antisense strand; or may have an RRMS at an internal position of the sense strand and an RRMS at an internal position of the antisense strand.

25 In preferred embodiments the iRNA agent includes a first and second sequences, which are preferably two separate molecules as opposed to two sequences located on the same strand, have sufficient complementarity to each other to hybridize (and thereby form a duplex region),

Attorney's Docket No.: 14174-072W01

e.g., under physiological conditions, e.g., under physiological conditions but not in contact with a helicase or other unwinding enzyme.

It is preferred that the first and second sequences be chosen such that the ds iRNA agent includes a single strand or unpaired region at one or both ends of the molecule. Thus, a ds iRNA agent contains first and second sequences, preferable paired to contain an overhang, e.g., one or two 5' or 3' overhangs but preferably a 3' overhang of 2-3 nucleotides. Most embodiments will have a 3' overhang. Preferred sRNA agents will have single-stranded overhangs, preferably 3' overhangs, of 1 or preferably 2 or 3 nucleotides in length at each end. The overhangs can be the result of one strand being longer than the other, or the result of two strands of the same length being staggered. 5' ends are preferably phosphorylated.

An RNA agent, e.g., an iRNA agent, containing a preferred, but nonlimiting RRMS is presented as formula (R-2) in FIG. 4. The carrier includes two "backbone attachment points" (hydroxyl groups), a "tethering attachment point," and a ligand, which is connected indirectly to the carrier via an intervening tether. The RRMS may be the 5' or 3' terminal subunit of the RNA molecule, i.e., one of the two "W" groups may be a hydroxyl group, and the other "W" group may be a chain of two or more unmodified or modified ribonucleotides. Alternatively, the RRMS may occupy an internal position, and both "W" groups may be one or more unmodified or modified ribonucleotides. More than one RRMS may be present in a RNA molecule, e.g., an iRNA agent.

The modified RNA molecule of formula (R-2) can be obtained using oligonucleotide synthetic methods known in the art. In a preferred embodiment, the modified RNA molecule of formula (II) can be prepared by incorporating one or more of the corresponding RRMS monomer compounds (RRMS monomers, see, e.g., A, B, and C in FIG. 4) into a growing sense or antisense strand, utilizing, e.g., phosphoramidite or H-phosphonate coupling strategies.

The RRMS monomers generally include two differently functionalized hydroxyl groups (OFG¹ and OFG² above), which are linked to the carrier molecule (see A in FIG. 4), and a tethering attachment point. As used herein, the term "functionalized hydroxyl group" means that the hydroxyl proton has been replaced by another substituent. As shown in representative structures B and C, one hydroxyl group (OFG¹) on the carrier is functionalized with a protecting

Attorney's Docket No.: 14174-072W01

group (PG). The other hydroxyl group (OFG²) can be functionalized with either (1) a liquid or solid phase synthesis support reagent (solid circle) directly or indirectly through a linker, L, as in B, or (2) a phosphorus-containing moiety, e.g., a phosphoramidite as in C. The tethering attachment point may be connected to a hydrogen atom, a tether, or a tethered ligand at the time that the monomer is incorporated into the growing sense or antisense strand (see R in Scheme 1). Thus, the tethered ligand can be, but need not be attached to the monomer at the time that the monomer is incorporated into the growing strand. In certain embodiments, the tether, the ligand or the tethered ligand may be linked to a "precursor" RRMS after a "precursor" RRMS monomer has been incorporated into the strand.

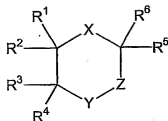
The (OFG¹) protecting group may be selected as desired, e.g., from T.W. Greene and P.G.M. Wuts, *Protective Groups in Organic Synthesis*, 2d. Ed., John Wiley and Sons (1991). The protecting group is preferably stable under amidite synthesis conditions, storage conditions, and oligonucleotide synthesis conditions. Hydroxyl groups, -OH, are nucleophilic groups (i.e., Lewis bases), which react through the oxygen with electrophiles (i.e., Lewis acids). Hydroxyl groups in which the hydrogen has been replaced with a protecting group, e.g., a triarylmethyl group or a trialkylsilyl group, are essentially unreactive as nucleophiles in displacement reactions. Thus, the protected hydroxyl group is useful in preventing e.g., homocoupling of compounds exemplified by structure C during oligonucleotide synthesis. A preferred protecting group is the dimethoxytrityl group.

When the OFG² in B includes a linker, e.g., a long organic linker, connected to a soluble or insoluble support reagent, solution or solid phase synthesis techniques can be employed to build up a chain of natural and/or modified ribonucleotides once OFG¹ is deprotected and free to react as a nucleophile with another nucleoside or monomer containing an electrophilic group (e.g., an amidite group). Alternatively, a natural or modified ribonucleotide or oligoribonucleotide chain can be coupled to monomer C via an amidite group or H-phosphonate group at OFG². Subsequent to this operation, OFG¹ can be deblocked, and the restored nucleophilic hydroxyl group can react with another nucleoside or monomer containing an electrophilic group (see FIG. 1). R' can be substituted or unsubstituted alkyl or alkenyl. In preferred embodiments, R' is methyl, allyl or 2-cyanoethyl. R'' may be a C₁-C₁₀ alkyl group, preferably it is a branched group containing three or more carbons, e.g., isopropyl.

Attorney's Docket No.: 14174-072W01

OFG² in **B** can be hydroxyl functionalized with a linker, which in turn contains a liquid or solid phase synthesis support reagent at the other linker terminus. The support reagent can be any support medium that can support the monomers described herein. The monomer can be attached to an insoluble support via a linker, L, which allows the monomer (and the growing chain) to be solubilized in the solvent in which the support is placed. The solubilized, yet immobilized, monomer can react with reagents in the surrounding solvent; unreacted reagents and soluble by-products can be readily washed away from the solid support to which the monomer or monomer-derived products is attached. Alternatively, the monomer can be attached to a soluble support moiety, e.g., polyethylene glycol (PEG) and liquid phase synthesis techniques can be used to build up the chain. Linker and support medium selection is within skill of the art. Generally the linker may be -C(O)(CH₂)₄C(O)-, or -C(O)(CH₂)₄S-, preferably, it is oxalyl, succinyl or thioglycolyl. Standard control pore glass solid phase synthesis supports can not be used in conjunction with fluoride labile 5' silyl protecting groups because the glass is degraded by fluoride with a significant reduction in the amount of full-length product. Fluoride-stable polystyrene based supports or PEG are preferred.

Preferred carriers have the general formula (R-3) provided below. (In that structure preferred backbone attachment points can be chosen from R¹ or R²; R³ or R⁴; or R⁹ and R¹⁰ if Y is CR⁹R¹⁰ (two positions are chosen to give two backbone attachment points, e.g., R¹ and R⁴, or R⁴ and R⁹. Preferred tethering attachment points include R⁷; R⁵ or R⁶ when X is CH₂. The carriers are described below as an entity, which can be incorporated into a strand. Thus, it is understood that the structures also encompass the situations wherein one (in the case of a terminal position) or two (in the case of an internal position) of the attachment points, e.g., R¹ or R²; R³ or R⁴; or R⁹ or R¹⁰ (when Y is CR⁹R¹⁰), is connected to the phosphate, or modified phosphate, e.g., sulfur containing, backbone. E.g., one of the above-named R groups can be -CH₂-, wherein one bond is connected to the carrier and one to a backbone atom, e.g., a linking oxygen or a central phosphorus atom.)



(R-3)

X is $\text{N}(\text{CO})\text{R}^7$, NR^7 or CH_2 ; Y is NR^8 , O, S, CR^9R^{10} ; and Z is $\text{CR}^{11}\text{R}^{12}$ or absent.

- 5 Each of R^1 , R^2 , R^3 , R^4 , R^9 , and R^{10} is, independently, H, OR^a , or $(\text{CH}_2)_n\text{OR}^b$, provided that at least two of R^1 , R^2 , R^3 , R^4 , R^9 , and R^{10} are OR^a and/or $(\text{CH}_2)_n\text{OR}^b$.

Each of R^5 , R^6 , R^{11} , and R^{12} is, independently, a ligand, H, $\text{C}_1\text{-C}_6$ alkyl optionally substituted with 1-3 R^{13} , or $\text{C}(\text{O})\text{NHR}^7$; or R^5 and R^{11} together are $\text{C}_3\text{-C}_8$ cycloalkyl optionally substituted with R^{14} .

- 10 R^7 is H, a ligand, or $\text{C}_1\text{-C}_{20}$ alkyl substituted with NR^cR^d ; R^8 is H or $\text{C}_1\text{-C}_6$ alkyl; R^{13} is hydroxy, $\text{C}_1\text{-C}_4$ alkoxy, or halo; R^{14} is NR^eR^7 ; R^{15} is $\text{C}_1\text{-C}_6$ alkyl optionally substituted with cyano, or $\text{C}_2\text{-C}_6$ alkenyl; R^{16} is $\text{C}_1\text{-C}_{10}$ alkyl; and R^{17} is a liquid or solid phase support reagent.

L is $-\text{C}(\text{O})(\text{CH}_2)_q\text{C}(\text{O})-$, or $-\text{C}(\text{O})(\text{CH}_2)_q\text{S}-$; R^a is CAr_3 ; R^b is $\text{P}(\text{O})(\text{O})\text{H}$, $\text{P}(\text{OR}^{15})\text{N}(\text{R}^{16})_2$ or L-R^{17} ; R^c is H or $\text{C}_1\text{-C}_6$ alkyl; and R^d is H or a ligand.

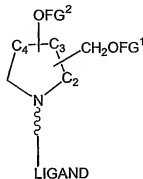
- 15 Each Ar is, independently, $\text{C}_6\text{-C}_{10}$ aryl optionally substituted with $\text{C}_1\text{-C}_4$ alkoxy; n is 1-4; and q is 0-4.

- Exemplary carriers include those in which, e.g., X is $\text{N}(\text{CO})\text{R}^7$ or NR^7 , Y is CR^9R^{10} , and Z is absent; or X is $\text{N}(\text{CO})\text{R}^7$ or NR^7 , Y is CR^9R^{10} , and Z is $\text{CR}^{11}\text{R}^{12}$; or X is $\text{N}(\text{CO})\text{R}^7$ or NR^7 , Y is NR^8 , and Z is $\text{CR}^{11}\text{R}^{12}$; or X is $\text{N}(\text{CO})\text{R}^7$ or NR^7 , Y is O, and Z is $\text{CR}^{11}\text{R}^{12}$; or X is CH_2 ; Y is CR^9R^{10} ; Z is $\text{CR}^{11}\text{R}^{12}$, and R^5 and R^{11} together form C_6 cycloalkyl (H, z = 2), or the indane ring
- 20

Attorney's Docket No.: 14174-072W01

system, e.g., X is CH₂; Y is CR⁹R¹⁰; Z is CR¹¹R¹², and R⁵ and R¹¹ together form C₅ cycloalkyl (H, z = 1).

- In certain embodiments, the carrier may be based on the pyrroline ring system or the 3-hydroxyproline ring system, e.g., X is N(CO)R⁷ or NR⁷, Y is CR⁹R¹⁰, and Z is absent (D). OFG¹ is preferably attached to a primary carbon, e.g., an exocyclic alkylene

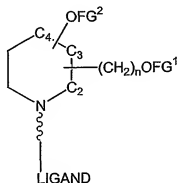


- group, e.g., a methylene group, connected to one of the carbons in the five-membered ring (-CH₂OFG¹ in D). OFG² is preferably attached directly to one of the carbons in the five-membered ring (-OFG² in D). For the pyrroline-based carriers, -CH₂OFG¹ may be attached to C-2 and OFG² may be attached to C-3; or -CH₂OFG¹ may be attached to C-3 and OFG² may be attached to C-4. . In certain embodiments, CH₂OFG¹ and OFG² may be geminally substituted to one of the above-referenced carbons. For the 3-hydroxyproline-based carriers, -CH₂OFG¹ may be attached to C-2 and OFG² may be attached to C-4. The pyrroline- and 3-hydroxyproline-based monomers may therefore contain linkages (e.g., carbon-carbon bonds) wherein bond rotation is restricted about that particular linkage, e.g. restriction resulting from the presence of a ring. Thus, CH₂OFG¹ and OFG² may be *cis* or *trans* with respect to one another in any of the pairings delineated above. Accordingly, all *cis/trans* isomers are expressly included. The monomers may also contain one or more asymmetric centers and thus occur as racemates and racemic mixtures, single enantiomers, individual diastereomers and diastereomeric mixtures. All such isomeric

Attorney's Docket No.: 14174-072W01

forms of the monomers are expressly included. The tethering attachment point is preferably nitrogen.

In certain embodiments, the carrier may be based on the piperidine ring system (E), e.g., X is $\text{N}(\text{CO})\text{R}^7$ or NR^7 , Y is CR^9R^{10} , and Z is $\text{CR}^{11}\text{R}^{12}$. OFG^1 is preferably



E

5

attached to a primary carbon, e.g., an exocyclic alkylene group, e.g., a methylene group ($n=1$) or ethylene group ($n=2$), connected to one of the carbons in the six-membered ring $[-(\text{CH}_2)_n\text{OFG}^1$ in E]. OFG^2 is preferably attached directly to one of the carbons in the six-membered ring ($-\text{OFG}^2$ in E). $-(\text{CH}_2)_n\text{OFG}^1$ and OFG^2 may be disposed in a geminal manner on the ring, i.e., both

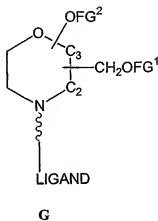
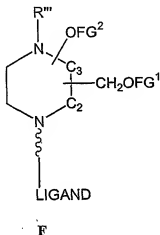
10 groups may be attached to the same carbon, e.g., at C-2, C-3, or C-4. Alternatively, $-(\text{CH}_2)_n\text{OFG}^1$ and OFG^2 may be disposed in a vicinal manner on the ring, i.e., both groups may be attached to adjacent ring carbon atoms, e.g., $-(\text{CH}_2)_n\text{OFG}^1$ may be attached to C-2 and OFG^2 may be attached to C-3; $-(\text{CH}_2)_n\text{OFG}^1$ may be attached to C-3 and OFG^2 may be attached to C-2; $-(\text{CH}_2)_n\text{OFG}^1$ may be attached to C-3 and OFG^2 may be attached to C-4; or $-(\text{CH}_2)_n\text{OFG}^1$ may be

15 attached to C-4 and OFG^2 may be attached to C-3. The piperidine-based monomers may therefore contain linkages (e.g., carbon-carbon bonds) wherein bond rotation is restricted about that particular linkage, e.g. restriction resulting from the presence of a ring. Thus, $-(\text{CH}_2)_n\text{OFG}^1$ and OFG^2 may be *cis* or *trans* with respect to one another in any of the pairings delineated above. Accordingly, all *cis/trans* isomers are expressly included. The monomers may also

Attorney's Docket No.: 14174-072W01

contain one or more asymmetric centers and thus occur as racemates and racemic mixtures, single enantiomers, individual diastereomers and diastereomeric mixtures. All such isomeric forms of the monomers are expressly included. The tethering attachment point is preferably nitrogen.

- 5 In certain embodiments, the carrier may be based on the piperazine ring system (F), e.g., X is $N(\text{CO})\text{R}^7$ or NR^7 , Y is NR^8 , and Z is $\text{CR}^{11}\text{R}^{12}$, or the morpholine ring system (G), e.g., X is $N(\text{CO})\text{R}^7$ or NR^7 , Y is O, and Z is $\text{CR}^{11}\text{R}^{12}$. OFG^1 is preferably

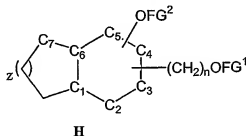


- attached to a primary carbon, e.g., an exocyclic alkylene group, e.g., a methylene group, connected to one of the carbons in the six-membered ring ($-\text{CH}_2\text{OFG}^1$ in F or G). OFG^2 is preferably attached directly to one of the carbons in the six-membered rings ($-\text{OFG}^2$ in F or G). For both F and G, $-\text{CH}_2\text{OFG}^1$ may be attached to C-2 and OFG^2 may be attached to C-3; or *vice versa*. In certain embodiments, CH_2OFG^1 and OFG^2 may be geminally substituted to one of the above-referenced carbons. The piperazine- and morpholine-based monomers may therefore contain linkages (e.g., carbon-carbon bonds) wherein bond rotation is restricted about that particular linkage, e.g. restriction resulting from the presence of a ring. Thus, CH_2OFG^1 and OFG^2 may be *cis* or *trans* with respect to one another in any of the pairings delineated above. Accordingly, all *cis/trans* isomers are expressly included. The monomers may also contain one or more asymmetric centers and thus occur as racemates and racemic mixtures, single

Attorney's Docket No.: 14174-072W01

enantiomers, individual diastereomers and diastereomeric mixtures. All such isomeric forms of the monomers are expressly included. R''' can be, e.g., C_1 - C_6 alkyl, preferably CH_3 . The tethering attachment point is preferably nitrogen in both **F** and **G**.

In certain embodiments, the carrier may be based on the decalin ring system, e.g., **X** is CH_2 ; **Y** is CR^9R^{10} ; **Z** is $CR^{11}R^{12}$, and R^5 and R^{11} together form C_6 cycloalkyl (**H**, $z = 2$), or the indane ring system, e.g., **X** is CH_2 ; **Y** is CR^9R^{10} ; **Z** is $CR^{11}R^{12}$, and R^5 and R^{11} together form C_5 cycloalkyl (**H**, $z = 1$). OFG^1 is preferably attached to a primary carbon,



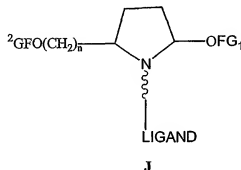
e.g., an exocyclic methylene group ($n=1$) or ethylene group ($n=2$) connected to one of C-2, C-3, C-4, or C-5 [$-(CH_2)_nOFG^1$ in **H**]. OFG^2 is preferably attached directly to one of C-2, C-3, C-4, or C-5 ($-OFG^2$ in **H**). $-(CH_2)_nOFG^1$ and OFG^2 may be disposed in a geminal manner on the ring, i.e., both groups may be attached to the same carbon, e.g., at C-2, C-3, C-4, or C-5.

Alternatively, $-(CH_2)_nOFG^1$ and OFG^2 may be disposed in a vicinal manner on the ring, i.e., both groups may be attached to adjacent ring carbon atoms, e.g., $-(CH_2)_nOFG^1$ may be attached to C-2 and OFG^2 may be attached to C-3; $-(CH_2)_nOFG^1$ may be attached to C-3 and OFG^2 may be attached to C-2; $-(CH_2)_nOFG^1$ may be attached to C-3 and OFG^2 may be attached to C-4; or $-(CH_2)_nOFG^1$ may be attached to C-4 and OFG^2 may be attached to C-3; $-(CH_2)_nOFG^1$ may be attached to C-4 and OFG^2 may be attached to C-5; or $-(CH_2)_nOFG^1$ may be attached to C-5 and OFG^2 may be attached to C-4. The decalin or indane-based monomers may therefore contain linkages (e.g., carbon-carbon bonds) wherein bond rotation is restricted about that particular linkage, e.g. restriction resulting from the presence of a ring. Thus, $-(CH_2)_nOFG^1$ and OFG^2 may be *cis* or *trans* with respect to one another in any of the pairings delineated above. Accordingly, all *cis/trans* isomers are expressly included. The monomers may also contain one or more

Attorney's Docket No.: 14174-072W01

asymmetric centers and thus occur as racemates and racemic mixtures, single enantiomers, individual diastereomers and diastereomeric mixtures. All such isomeric forms of the monomers are expressly included. In a preferred embodiment, the substituents at C-1 and C-6 are *trans* with respect to one another. The tethering attachment point is preferably C-6 or C-7.

- 5 Other carriers may include those based on 3-hydroxyproline (J). Thus, $-(CH_2)_nOFG^1$ and OFG^2 may be *cis* or *trans* with respect to one another. Accordingly, all *cis/trans* isomers are expressly included. The monomers may also contain one or more asymmetric centers



- and thus occur as racemates and racemic mixtures, single enantiomers, individual diastereomers and diastereomeric mixtures. All such isomeric forms of the monomers are expressly included. The tethering attachment point is preferably nitrogen.

Representative carriers are shown in FIG. 5.

- In certain embodiments, a moiety, e.g., a ligand may be connected indirectly to the carrier via the intermediacy of an intervening tether. Tethers are connected to the carrier at the tethering attachment point (TAP) and may include any C_1 - C_{100} carbon-containing moiety, (e.g. C_1 - C_{75} , C_1 - C_{50} , C_1 - C_{20} , C_1 - C_{10} , C_1 - C_6), preferably having at least one nitrogen atom. In preferred embodiments, the nitrogen atom forms part of a terminal amino group on the tether, which may serve as a connection point for the ligand. Preferred tethers (underlined) include $TAP-(CH_2)_nNH_2$; $TAP-C(O)(CH_2)_nNH_2$; or $TAP-NR^{''''}(CH_2)_nNH_2$, in which n is 1-6 and $R^{''''}$ is C_1 - C_6 alkyl and R^d is hydrogen or a ligand. In other embodiments, the nitrogen may form part of a terminal oxyamino group, e.g., $-ONH_2$, or hydrazino group, $-NHNH_2$. The tether may optionally

Attorney's Docket No.: 14174-072W01

be substituted, e.g., with hydroxy, alkoxy, perhaloalkyl, and/or optionally inserted with one or more additional heteroatoms, e.g., N, O, or S. Preferred tethered ligands may include, e.g., TAP-(CH₂)_nNH(LIGAND),

TAP-C(O)(CH₂)_nNH(LIGAND), or TAP-NR^{''''}(CH₂)_nNH(LIGAND);

5 TAP-(CH₂)_nONH(LIGAND), TAP-C(O)(CH₂)_nONH(LIGAND), or

TAP-NR^{''''}(CH₂)_nONH(LIGAND); TAP-(CH₂)_nNHNH₂(LIGAND),

TAP-C(O)(CH₂)_nNHNH₂(LIGAND), or TAP-NR^{''''}(CH₂)_nNHNH₂(LIGAND).

In other embodiments the tether may include an electrophilic moiety, preferably at the terminal position of the tether. Preferred electrophilic moieties include, e.g., an aldehyde, alkyl
10 halide, mesylate, tosylate, nosylate, or brosylate, or an activated carboxylic acid ester, e.g. an NHS ester, or a pentafluorophenyl ester. Preferred tethers (underlined) include TAP-(CH₂)_nCHO; TAP-C(O)(CH₂)_nCHO; or TAP-NR^{''''}(CH₂)_nCHO, in which n is 1-6 and R^{''''} is C₁-C₆ alkyl; or TAP-(CH₂)_nC(O)ONHS; TAP-C(O)(CH₂)_nC(O)ONHS; or
TAP-NR^{''''}(CH₂)_nC(O)ONHS, in which n is 1-6 and R^{''''} is C₁-C₆ alkyl;

15 TAP-(CH₂)_nC(O)OC₆F₅; TAP-C(O)(CH₂)_nC(O)OC₆F₅; or TAP-NR^{''''}(CH₂)_nC(O)OC₆F₅, in which n is 1-6 and R^{''''} is C₁-C₆ alkyl; or (CH₂)_nCH₂LG; TAP-C(O)(CH₂)_nCH₂LG; or TAP-NR^{''''}(CH₂)_nCH₂LG, in which n is 1-6 and R^{''''} is C₁-C₆ alkyl (LG can be a leaving group, e.g., halide, mesylate, tosylate, nosylate, brosylate). Tethering can be carried out by coupling a nucleophilic group of a ligand, e.g., a thiol or amino group with an electrophilic group on the
20 tether.

Tethered Entities

A wide variety of entities can be tethered to an iRNA agent, e.g., to the carrier of an RRMS. Examples are described below in the context of an RRMS but that is only preferred, entities can be coupled at other points to an iRNA agent. Preferred entities are those which
25 target to the liver.

Attorney's Docket No.: 14174-072W01

Preferred moieties are ligands, which are coupled, preferably covalently, either directly or indirectly via an intervening tether, to the RRMS carrier. In preferred embodiments, the ligand is attached to the carrier *via* an intervening tether. As discussed above, the ligand or tethered ligand may be present on the RRMS monomer when the RRMS monomer is incorporated into the growing strand. In some embodiments, the ligand may be incorporated into a "precursor" RRMS after a "precursor" RRMS monomer has been incorporated into the growing strand. For example, an RRMS monomer having, e.g., an amino-terminated tether (i.e., having no associated ligand), e.g., TAP-(CH₂)_nNH₂ may be incorporated into a growing sense or antisense strand. In a subsequent operation, i.e., after incorporation of the precursor monomer into the strand, a ligand having an electrophilic group, e.g., a pentafluorophenyl ester or aldehyde group, can subsequently be attached to the precursor RRMS by coupling the electrophilic group of the ligand with the terminal nucleophilic group of the precursor RRMS tether.

In preferred embodiments, a ligand alters the distribution, targeting or lifetime of an iRNA agent into which it is incorporated. In preferred embodiments a ligand provides an enhanced affinity for a selected target, e.g. molecule, cell or cell type, compartment, e.g., a cellular or organ compartment, tissue, organ or region of the body, as, e.g., compared to a species absent such a ligand. Preferred ligands will not take part in duplex pairing in a duplexed nucleic acid.

Preferred ligands can improve transport, hybridization, and specificity properties and may also improve nuclease resistance of the resultant natural or modified oligoribonucleotide, or a polymeric molecule comprising any combination of monomers described herein and/or natural or modified ribonucleotides.

Ligands in general can include therapeutic modifiers, e.g., for enhancing uptake; diagnostic compounds or reporter groups e.g., for monitoring distribution; cross-linking agents; and nuclease-resistance conferring moieties. General examples include lipids, steroids, vitamins, sugars, proteins, peptides, polyamines, and peptide mimics.

Ligands can include a naturally occurring substance, such as a protein (e.g., human serum albumin (HSA), low-density lipoprotein (LDL), or globulin); carbohydrate (e.g., a dextran, pullulan, chitin, chitosan, inulin, cyclodextrin or hyaluronic acid); or a lipid. The ligand may

Attorney's Docket No.: 14174-072W01

also be a recombinant or synthetic molecule, such as a synthetic polymer, e.g., a synthetic polyamino acid. Examples of polyamino acids include polyamino acid is a polylysine (PLL), poly L-aspartic acid, poly L-glutamic acid, styrene-maleic acid anhydride copolymer, poly(L-lactide-co-glycolid) copolymer, divinyl ether-maleic anhydride copolymer, N-(2-hydroxypropyl)methacrylamide copolymer (HMPA), polyethylene glycol (PEG), polyvinyl alcohol (PVA), polyurethane, poly(2-ethylacrylic acid), N-isopropylacrylamide polymers, or polyphosphazine. Example of polyamines include: polyethylenimine, polylysine (PLL), spermine, spermidine, polyamine, pseudopeptide-polyamine, peptidomimetic polyamine, dendrimer polyamine, arginine, amidine, protamine, cationic lipid, cationic porphyrin, quaternary salt of a polyamine, or an alpha helical peptide.

Ligands can also include targeting groups, e.g., a cell or tissue targeting agent, e.g., a lectin, glycoprotein, lipid or protein, e.g., an antibody, that binds to a specified cell type such as a liver cell. A targeting group can be a thyrotropin, melanotropin, lectin, glycoprotein, surfactant protein A, Mucin carbohydrate, multivalent lactose, multivalent galactose, N-acetyl-galactosamine, N-acetyl-gulucosamine multivalent mannose, multivalent fucose, glycosylated polyaminoacids, multivalent galactose, transferrin, bisphosphonate, polyglutamate, polyaspartate, a lipid, cholesterol, a steroid, bile acid, folate, vitamin B12, biotin, or an RGD peptide or RGD peptide mimetic.

Other examples of ligands include dyes, intercalating agents (e.g. acridines), cross-linkers (e.g. psoralene, mitomycin C), porphyrins (TPPC4, texaphyrin, Sapphyrin), polycyclic aromatic hydrocarbons (e.g., phenazine, dihydrophenazine), artificial endonucleases (e.g. EDTA), lipophilic molecules, e.g. cholesterol, cholic acid, adamantane acetic acid, 1-pyrene butyric acid, dihydrotestosterone, 1,3-Bis-O(hexadecyl)glycerol, geranyloxyhexyl group, hexadecylglycerol, borneol, menthol, 1,3-propanediol, heptadecyl group, palmitic acid, myristic acid, O3-(oleoyl)lithocholic acid, O3-(oleoyl)cholenic acid, dimethoxytrityl, or phenoxazine) and peptide conjugates (e.g., antennapedia peptide, Tat peptide), alkylating agents, phosphate, amino, mercapto, PEG (e.g., PEG-40K), MPEG, [MPEG]₂, polyamino, alkyl, substituted alkyl, radiolabeled markers, enzymes, haptens (e.g. biotin), transport/absorption facilitators (e.g., aspirin, vitamin E, folic acid), synthetic ribonucleases (e.g., imidazole, bisimidazole, histamine,

Attorney's Docket No.: 14174-072W01

imidazole clusters, acridine-imidazole conjugates, Eu3+ complexes of tetraazamacrocycles), dinitrophenyl, HRP, or AP.

5 Ligands can be proteins, e.g., glycoproteins, or peptides, e.g., molecules having a specific affinity for a co-ligand, or antibodies e.g., an antibody, that binds to a specified cell type such as a cancer cell, endothelial cell, or bone cell. Ligands may also include hormones and hormone receptors. They can also include non-peptidic species, such as lipids, lectins, carbohydrates, vitamins, cofactors, multivalent lactose, multivalent galactose, N-acetyl-galactosamine, N-acetyl-gulucosamine multivalent mannose, or multivalent fucose. The ligand can be, for example, a lipopolysaccharide, an activator of p38 MAP kinase, or an activator of NF- κ B.

10 The ligand can be a substance, e.g., a drug, which can increase the uptake of the iRNA agent into the cell, for example, by disrupting the cell's cytoskeleton, e.g., by disrupting the cell's microtubules, microfilaments, and/or intermediate filaments. The drug can be, for example, taxon, vincristine, vinblastine, cytochalasin, nocodazole, japlakinolide, latrunculin A, phalloidin, swinholide A, indanocine, or myoservin.

15 The ligand can increase the uptake of the iRNA agent into the cell by activating an inflammatory response, for example. Exemplary ligands that would have such an effect include tumor necrosis factor alpha (TNFalpha), interleukin-1 beta, or gamma interferon.

20 In one aspect, the ligand is a lipid or lipid-based molecule. Such a lipid or lipid-based molecule preferably binds a serum protein, e.g., human serum albumin (HSA). An HSA binding ligand allows for distribution of the conjugate to a target tissue, e.g., a non-liver target tissue of the body. Preferably, the target tissue is the liver, preferably parenchymal cells of the liver. Other molecules that can bind HSA can also be used as ligands. For example, neproxin or aspirin can be used. A lipid or lipid-based ligand can (a) increase resistance to degradation of the conjugate, (b) increase targeting or transport into a target cell or cell membrane, and/or (c) can be
25 used to adjust binding to a serum protein, e.g., HSA.

A lipid based ligand can be used to modulate, e.g., control the binding of the conjugate to a target tissue. For example, a lipid or lipid-based ligand that binds to HSA more strongly will be less likely to be targeted to the liver and therefore less likely to be cleared from the body.

Attorney's Docket No.: 14174-072W01

In a preferred embodiment, the lipid based ligand binds HSA. Preferably, it binds HSA with a sufficient affinity such that the conjugate will be preferably distributed to a non-kidney tissue. However, it is preferred that the affinity not be so strong that the HSA-ligand binding cannot be reversed.

5 In another aspect, the ligand is a moiety, e.g., a vitamin, which is taken up by a target cell, e.g., a proliferating cell. These are particularly useful for treating disorders characterized by unwanted cell proliferation, e.g., of the malignant or non-malignant type, e.g., cancer cells. Exemplary vitamins include vitamin A, E, and K. Other exemplary vitamins include are B vitamin, e.g., folic acid, B12, riboflavin, biotin, pyridoxal or other vitamins or nutrients taken up
10 by cancer cells. Also included are HSA and low density lipoprotein (LDL).

In another aspect, the ligand is a cell-permeation agent, preferably a helical cell-permeation agent. Preferably, the agent is amphipathic. An exemplary agent is a peptide such as tat or antennopodia. If the agent is a peptide, it can be modified, including a peptidylmimetic, invertomers, non-peptide or pseudo-peptide linkages, and use of D-amino acids. The helical
15 agent is preferably an alpha-helical agent, which preferably has a lipophilic and a lipophobic phase.

The ligand can be a peptide or peptidomimetic. A peptidomimetic (also referred to herein as an oligopeptidomimetic) is a molecule capable of folding into a defined three-dimensional structure similar to a natural peptide. The attachment of peptide and
20 peptidomimetics to iRNA agents can affect pharmacokinetic distribution of the iRNA, such as by enhancing cellular recognition and absorption. The peptide or peptidomimetic moiety can be about 5-50 amino acids long, e.g., about 5, 10, 15, 20, 25, 30, 35, 40, 45, or 50 amino acids long (see Table 2, for example).

Table 2. Exemplary Cell Permeation Peptides

Cell Permeation Peptide	Amino acid Sequence	Reference
Penetratin	RQIKIWFQNRRMKWKK (SEQ ID NO:6700)	Derossi <i>et al.</i> , J. Biol. Chem. 269:10444, 1994
Tat fragment (48-60)	GRKKRRQRRPPQC (SEQ ID NO:6701)	Vives <i>et al.</i> , J. Biol. Chem., 272:16010, 1997
Signal Sequence-based peptide	GALFLGWLGAAGSTMGAWSQPKKKRKV (SEQ ID NO:6702)	Chaloin <i>et al.</i> , Biochem. Biophys. Res. Commun., 243:601, 1998
PVEC	LLIILRRIRKQAHASK (SEQ ID NO:6703)	Elmqvist <i>et al.</i> , Exp. Cell Res., 269:237, 2001
Transportan	GWTLSAGYLLKINLKALAALAKKIL (SEQ ID NO:6704)	Pooga <i>et al.</i> , FASEB J., 12:67, 1998

Amphiphilic model peptide	KLALKLALKALKAAALKLA (SEQ ID NO:6705)	Oehlke <i>et al.</i> , Mol. Ther., 2:339, 2000
Arg ₉	RRRRRRRRR (SEQ ID NO:6706)	Mitchell <i>et al.</i> , J. Pept. Res., 56:318, 2000
Bacterial cell wall permeating	KFFKFFKFFK (SEQ ID NO:6707)	
LL-37	LLGDFFRKSKEKIGKEFKRIVQRIKDFLRN LVPRTES (SEQ ID NO:6708)	
Cecropin P1	SWLSKTAKKLENSAKKRRISEGIALAIQGGP R (SEQ ID NO:6709)	
α -defensin	ACYCRIPACIAGERRYGTCTYQGRLWAFCC (SEQ ID NO:6710)	
b-defensin	DHYNCVSSGGQCLYSACPIFTKIQTGTCYR GKAKCCK (SEQ ID NO:6711)	
Bactenecin	RKCRIVVIRVCR (SEQ ID NO:6712)	
PR-39	RRRPRPPYLPRPRPPFFPPRLPPRIPPGFPP RFPFRFPGKR-NH ₂ (SEQ ID NO:6713)	
Indolicidin	ILPWKWPWWPWRR-NH ₂ (SEQ ID NO:6714)	

A peptide or peptidomimetic can be, for example, a cell permeation peptide, cationic peptide, amphipathic peptide, or hydrophobic peptide (e.g., consisting primarily of Tyr, Trp or Phe). The peptide moiety can be a dendrimer peptide, constrained peptide or crosslinked peptide. The peptide moiety can be an L-peptide or D-peptide. In another alternative, the

Attorney's Docket No.: 14174-072W01

peptide moiety can include a hydrophobic membrane translocation sequence (MTS). An exemplary hydrophobic MTS-containing peptide is RFGF having the amino acid sequence AAVALLPAVLLALLAP (SEQ ID NO:6715). An RFGF analogue (*e.g.*, amino acid sequence AALLPVLLAAP (SEQ ID NO:6716)) containing a hydrophobic MTS can also be a targeting moiety. The peptide moiety can be a "delivery" peptide, which can carry large polar molecules including peptides, oligonucleotides, and protein across cell membranes. For example, sequences from the HIV Tat protein (GRKKRRQRRPPQ (SEQ ID NO:6717)) and the Drosophila Antennapedia protein (RQIKIWFQNRRMKWKK (SEQ ID NO:6718)) have been found to be capable of functioning as delivery peptides. A peptide or peptidomimetic can be encoded by a random sequence of DNA, such as a peptide identified from a phage-display library, or one-bead-one-compound (OBOC) combinatorial library (Lam *et al.*, Nature, 354:82-84, 1991). Preferably the peptide or peptidomimetic tethered to an iRNA agent via an incorporated monomer unit is a cell targeting peptide such as an arginine-glycine-aspartic acid (RGD)-peptide, or RGD mimic. A peptide moiety can range in length from about 5 amino acids to about 40 amino acids. The peptide moieties can have a structural modification, such as to increase stability or direct conformational properties. Any of the structural modifications described below can be utilized.

An RGD peptide moiety can be used to target a tumor cell, such as an endothelial tumor cell or a breast cancer tumor cell (Zitzmann *et al.*, Cancer Res., 62:5139-43, 2002). An RGD peptide can facilitate targeting of an iRNA agent to tumors of a variety of other tissues, including the lung, kidney, spleen, or liver (Aoki *et al.*, Cancer Gene Therapy 8:783-787, 2001). The RGD peptide can be linear or cyclic, and can be modified, *e.g.*, glycosylated or methylated to facilitate targeting to specific tissues. For example, a glycosylated RGD peptide can deliver an iRNA agent to a tumor cell expressing $\alpha_v\beta_3$ (Haubner *et al.*, Jour. Nucl. Med., 42:326-336, 2001).

Peptides that target markers enriched in proliferating cells can be used. *E.g.*, RGD containing peptides and peptidomimetics can target cancer cells, in particular cells that exhibit an $\alpha_v\beta_3$ integrin. Thus, one could use RGD peptides, cyclic peptides containing RGD, RGD peptides that include D-amino acids, as well as synthetic RGD mimics. In addition to RGD, one can use other moieties that target the $\alpha_v\beta_3$ integrin ligand. Generally, such ligands can be used

Attorney's Docket No.: 14174-072W01

to control proliferating cells and angiogenesis. Preferred conjugates of this type include an iRNA agent that targets PECAM-1, VEGF, or other cancer gene, e.g., a cancer gene described herein.

A "cell permeation peptide" is capable of permeating a cell, e.g., a microbial cell, such as a bacterial or fungal cell, or a mammalian cell, such as a human cell. A microbial cell-permeating peptide can be, for example, an α -helical linear peptide (e.g., LL-37 or Ceropin P1), a disulfide bond-containing peptide (e.g., α -defensin, β -defensin or bactenecin), or a peptide containing only one or two dominating amino acids (e.g., PR-39 or indolicidin). A cell permeation peptide can also include a nuclear localization signal (NLS). For example, a cell permeation peptide can be a bipartite amphipathic peptide, such as MPG, which is derived from the fusion peptide domain of HIV-1 gp41 and the NLS of SV40 large T antigen (Simeoni *et al.*, Nucl. Acids Res. 31:2717-2724, 2003).

In one embodiment, a targeting peptide tethered to an RRMS can be an amphipathic α -helical peptide. Exemplary amphipathic α -helical peptides include, but are not limited to, cecropins, lycotoxins, paradaxins, buforin, CPF, bombinin-like peptide (BLP), cathelicidins, ceratotoxins, *S. clava* peptides, hagfish intestinal antimicrobial peptides (HFIAPs), magainines, brevinins-2, dermaseptins, melittins, pleurocidin, H₂A peptides, *Xenopus* peptides, esculentinisin-1, and caerins. A number of factors will preferably be considered to maintain the integrity of helix stability. For example, a maximum number of helix stabilization residues will be utilized (e.g., leu, ala, or lys), and a minimum number helix destabilization residues will be utilized (e.g., proline, or cyclic monomeric units. The capping residue will be considered (for example Gly is an exemplary N-capping residue and/or C-terminal amidation can be used to provide an extra H-bond to stabilize the helix. Formation of salt bridges between residues with opposite charges, separated by $i \pm 3$, or $i \pm 4$ positions can provide stability. For example, cationic residues such as lysine, arginine, homo-arginine, ornithine or histidine can form salt bridges with the anionic residues glutamate or aspartate.

Peptide and petidomimetic ligands include those having naturally occurring or modified peptides, e.g., D or L peptides; α , β , or γ peptides; N-methyl peptides; azapeptides; peptides having one or more amide, i.e., peptide, linkages replaced with one or more urea, thiourea, carbamate, or sulfonyl urea linkages; or cyclic peptides.

Attorney's Docket No.: 14174-072W01

Methods for making iRNA agents

iRNA agents can include modified or non-naturally occurring bases, e.g., bases described in copending and coowned United States Provisional Application Serial No. 60/463,772, filed on April 17, 2003, which is hereby incorporated by reference and/or in copending and coowned United States Provisional Application Serial No. 60/465,802, filed on April 25, 2003, which is hereby incorporated by reference. Monomers and iRNA agents which include such bases can be made by the methods found in United States Provisional Application Serial No. 60/463,772, filed on April 17, 2003, and/or in United States Provisional Application Serial No. 60/465,802, filed on April 25, 2003.

In addition, the invention includes iRNA agents having a modified or non-naturally occurring base and another element described herein. E.g., the invention includes an iRNA agent described herein, e.g., a palindromic iRNA agent, an iRNA agent having a non canonical pairing, an iRNA agent which targets a gene described herein, e.g., a gene active in the liver, an iRNA agent having an architecture or structure described herein, an iRNA associated with an amphipathic delivery agent described herein, an iRNA associated with a drug delivery module described herein, an iRNA agent administered as described herein, or an iRNA agent formulated as described herein, which also incorporates a modified or non-naturally occurring base.

The synthesis and purification of oligonucleotide peptide conjugates can be performed by established methods. See, for example, Trufert *et al.*, *Tetrahedron*, 52:3005, 1996; and Manoharan, "Oligonucleotide Conjugates in Antisense Technology," in Antisense Drug Technology, ed. S.T. Crooke, Marcel Dekker, Inc., 2001.

In one embodiment of the invention, a peptidomimetic can be modified to create a constrained peptide that adopts a distinct and specific preferred conformation, which can increase the potency and selectivity of the peptide. For example, the constrained peptide can be an azapeptide (Gante, *Synthesis*, 405-413, 1989). An azapeptide is synthesized by replacing the α -carbon of an amino acid with a nitrogen atom without changing the structure of the amino acid side chain. For example, the azapeptide can be synthesized by using hydrazine in traditional peptide synthesis coupling methods, such as by reacting hydrazine with a "carbonyl donor," e.g., phenylchloroformate.

Attorney's Docket No.: 14174-072W01

In one embodiment of the invention, a peptide or peptidomimetic (*e.g.*, a peptide or peptidomimetic tethered to an RRMS) can be an N-methyl peptide. N-methyl peptides are composed of N-methyl amino acids, which provide an additional methyl group in the peptide backbone, thereby potentially providing additional means of resistance to proteolytic cleavage.

- 5 N-methyl peptides can be synthesized by methods known in the art (see, for example, Lindgren *et al.*, Trends Pharmacol. Sci. 21:99, 2000; Cell Penetrating Peptides: Processes and Applications, Langel, ed., CRC Press, Boca Raton, FL, 2002; Fische *et al.*, Bioconjugate. Chem. 12: 825, 2001; Wander *et al.*, J. Am. Chem. Soc., 124:13382, 2002). For example, an Ant or Tat peptide can be an N-methyl peptide.

- 10 In one embodiment of the invention, a peptide or peptidomimetic (*e.g.*, a peptide or peptidomimetic tethered to an RRMS) can be a β -peptide. β -peptides form stable secondary structures such as helices, pleated sheets, turns and hairpins in solutions. Their cyclic derivatives can fold into nanotubes in the solid state. β -peptides are resistant to degradation by proteolytic enzymes. β -peptides can be synthesized by methods known in the art. For example, an Ant or
- 15 Tat peptide can be a β -peptide.

In one embodiment of the invention, a peptide or peptidomimetic (*e.g.*, a peptide or peptidomimetic tethered to an RRMS) can be an oligocarbamate. Oligocarbamate peptides are internalized into a cell by a transport pathway facilitated by carbamate transporters. For example, an Ant or Tat peptide can be an oligocarbamate.

- 20 In one embodiment of the invention, a peptide or peptidomimetic (*e.g.*, a peptide or peptidomimetic tethered to an RRMS) can be an oligourea conjugate (or an oligothiourea conjugate), in which the amide bond of a peptidomimetic is replaced with a urea moiety. Replacement of the amide bond provides increased resistance to degradation by proteolytic enzymes, *e.g.*, proteolytic enzymes in the gastrointestinal tract. In one embodiment, an oligourea
- 25 conjugate is tethered to an iRNA agent for use in oral delivery. The backbone in each repeating unit of an oligourea peptidomimetic can be extended by one carbon atom in comparison with the natural amino acid. The single carbon atom extension can increase peptide stability and lipophilicity, for example. An oligourea peptide can therefore be advantageous when an iRNA agent is directed for passage through a bacterial cell wall, or when an iRNA agent must traverse

Attorney's Docket No.: 14174-072W01

the blood-brain barrier, such as for the treatment of a neurological disorder. In one embodiment, a hydrogen bonding unit is conjugated to the oligoureia peptide, such as to create an increased affinity with a receptor. For example, an Ant or Tat peptide can be an oligoureia conjugate (or an oligothioureia conjugate).

5 The siRNA peptide conjugates of the invention can be affiliated with, *e.g.*, tethered to, RRMSs occurring at various positions on an iRNA agent. For example, a peptide can be terminally conjugated, on either the sense or the antisense strand, or a peptide can be bisconjugated (one peptide tethered to each end, one conjugated to the sense strand, and one conjugated to the antisense strand). In another option, the peptide can be internally conjugated,
10 such as in the loop of a short hairpin iRNA agent. In yet another option, the peptide can be affiliated with a complex, such as a peptide-carrier complex.

 A peptide-carrier complex consists of at least a carrier molecule, which can encapsulate one or more iRNA agents (such as for delivery to a biological system and/or a cell), and a peptide moiety tethered to the outside of the carrier molecule, such as for targeting the carrier
15 complex to a particular tissue or cell type. A carrier complex can carry additional targeting molecules on the exterior of the complex, or fusogenic agents to aid in cell delivery. The one or more iRNA agents encapsulated within the carrier can be conjugated to lipophilic molecules, which can aid in the delivery of the agents to the interior of the carrier.

 A carrier molecule or structure can be, for example, a micelle, a liposome (*e.g.*, a cationic
20 liposome), a nanoparticle, a microsphere, or a biodegradable polymer. A peptide moiety can be tethered to the carrier molecule by a variety of linkages, such as a disulfide linkage, an acid labile linkage, a peptide-based linkage, an oxyamino linkage or a hydrazine linkage. For example, a peptide-based linkage can be a GFLG peptide. Certain linkages will have particular advantages, and the advantages (or disadvantages) can be considered depending on the tissue
25 target or intended use. For example, peptide based linkages are stable in the blood stream but are susceptible to enzymatic cleavage in the lysosomes.

Attorney's Docket No.: 14174-072W01

Targeting

The iRNA agents of the invention are particularly useful when targeted to the liver. An iRNA agent can be targeted to the liver by incorporation of an RRMS containing a ligand that targets the liver. For example, a liver-targeting agent can be a lipophilic moiety. Preferred lipophilic moieties include lipids, cholesterol, oleyl, retinyl, or cholesterol residues. Other lipophilic moieties that can function as liver-targeting agents include cholic acid, adamantane acetic acid, 1-pyrene butyric acid, dihydrotestosterone, 1,3-Bis-O(hexadecyl)glycerol, geranyloxyhexyl group, hexadecylglycerol, borneol, menthol, 1,3-propanediol, heptadecyl group, palmitic acid, myristic acid, O3-(oleoyl)lithocholic acid, O3-(oleoyl)cholenic acid, dimethoxytrityl, or phenoxazine.

An iRNA agent can also be targeted to the liver by association with a low-density lipoprotein (LDL), such as lactosylated LDL. Polymeric carriers complexed with sugar residues can also function to target iRNA agents to the liver.

A targeting agent that incorporates a sugar, *e.g.*, galactose and/or analogues thereof, is particularly useful. These agents target, in particular, the parenchymal cells of the liver. For example, a targeting moiety can include more than one or preferably two or three galactose moieties, spaced about 15 angstroms from each other. The targeting moiety can alternatively be lactose (*e.g.*, three lactose moieties), which is glucose coupled to a galactose. The targeting moiety can also be N-Acetyl-Galactosamine, N-Ac-Glucosamine. A mannose or mannose-6-phosphate targeting moiety can be used for macrophage targeting.

Conjugation of an iRNA agent with a serum albumin (SA), such as human serum albumin, can also be used to target the iRNA agent to a non-kidney tissue, such as the liver.

An iRNA agent targeted to the liver by an RRMS targeting moiety described herein can target a gene expressed in the liver.

An iRNA agent targeted to the liver by an RRMS targeting moiety described herein can target a gene expressed in the liver. For example, the iRNA agent can target p21(WAF1/DIP1), P27(KIP1), the α -fetoprotein gene, beta-catenin, or c-MET, such as for treating a cancer of the liver. In another embodiment, the iRNA agent can target apoB-100, such as for the treatment of

Attorney's Docket No.: 14174-072W01

an HDL/LDL cholesterol imbalance; dyslipidemias, e.g., familial combined hyperlipidemia (FCHL), or acquired hyperlipidemia; hypercholesterolemia; statin-resistant hypercholesterolemia; coronary artery disease (CAD); coronary heart disease (CHD); or atherosclerosis. In another embodiment, the iRNA agent can target forkhead homologue in rhabdomyosarcoma (FKHR); glucagon; glucagon receptor; glycogen phosphorylase; PPAR-Gamma Coactivator (PGC-1); fructose-1,6-bisphosphatase; glucose-6-phosphatase; glucose-6-phosphate translocator; glucokinase inhibitory regulatory protein; or phosphoenolpyruvate carboxykinase (PEPCK), such as to inhibit hepatic glucose production in a mammal, such as a human, such as for the treatment of diabetes. In another embodiment, an iRNA agent targeted to the liver can target Factor V, e.g., the Leiden Factor V allele, such as to reduce the tendency to form a blood clot. An iRNA agent targeted to the liver can include a sequence which targets hepatitis virus (e.g., Hepatitis A, B, C, D, E, F, G, or H). For example, an iRNA agent of the invention can target any one of the nonstructural proteins of HCV: NS3, 4A, 4B, 5A, or 5B. For the treatment of hepatitis B, an iRNA agent can target the protein X (HBx) gene, for example.

Preferred ligands on RRMSSs include folic acid, glucose, cholesterol, cholic acid, Vitamin E, Vitamin K, or Vitamin A.

Definitions

The term "halo" refers to any radical of fluorine, chlorine, bromine or iodine.

The term "alkyl" refers to a hydrocarbon chain that may be a straight chain or branched chain, containing the indicated number of carbon atoms. For example, C₁-C₁₂ alkyl indicates that the group may have from 1 to 12 (inclusive) carbon atoms in it. The term "haloalkyl" refers to an alkyl in which one or more hydrogen atoms are replaced by halo, and includes alkyl moieties in which all hydrogens have been replaced by halo (e.g., perfluoroalkyl). Alkyl and haloalkyl groups may be optionally inserted with O, N, or S. The terms "aralkyl" refers to an alkyl moiety in which an alkyl hydrogen atom is replaced by an aryl group. Aralkyl includes groups in which more than one hydrogen atom has been replaced by an aryl group. Examples of "aralkyl" include benzyl, 9-fluorenyl, benzhydryl, and trityl groups.

Attorney's Docket No.: 14174-072W01

The term "alkenyl" refers to a straight or branched hydrocarbon chain containing 2-8 carbon atoms and characterized in having one or more double bonds. Examples of a typical alkenyl include, but not limited to, allyl, propenyl, 2-butenyl, 3-hexenyl and 3-octenyl groups. The term "alkynyl" refers to a straight or branched hydrocarbon chain containing 2-8 carbon atoms and characterized in having one or more triple bonds. Some examples of a typical alkynyl are ethynyl, 2-propynyl, and 3-methylbutynyl, and propargyl. The sp^2 and sp^3 carbons may optionally serve as the point of attachment of the alkenyl and alkynyl groups, respectively.

The term "alkoxy" refers to an -O-alkyl radical. The term "aminoalkyl" refers to an alkyl substituted with an amino group. The term "mercapto" refers to an -SH radical. The term "thioalkoxy" refers to an -S-alkyl radical.

The term "alkylene" refers to a divalent alkyl (*i.e.*, -R-), e.g., -CH₂-, -CH₂CH₂-, and -CH₂CH₂CH₂-. The term "alkylenedioxy" refers to a divalent species of the structure -O-R-O-, in which R represents an alkylene.

The term "aryl" refers to an aromatic monocyclic, bicyclic, or tricyclic hydrocarbon ring system, wherein any ring atom capable of substitution can be substituted by a substituent. Examples of aryl moieties include, but are not limited to, phenyl, naphthyl, and anthracenyl.

The term "cycloalkyl" as employed herein includes saturated cyclic, bicyclic, tricyclic, or polycyclic hydrocarbon groups having 3 to 12 carbons, wherein any ring atom capable of substitution can be substituted by a substituent. The cycloalkyl groups herein described may also contain fused rings. Fused rings are rings that share a common carbon-carbon bond. Examples of cycloalkyl moieties include, but are not limited to, cyclohexyl, adamantyl, and norbornyl.

The term "heterocyclyl" refers to a nonaromatic 3-10 membered monocyclic, 8-12 membered bicyclic, or 11-14 membered tricyclic ring system having 1-3 heteroatoms if monocyclic, 1-6 heteroatoms if bicyclic, or 1-9 heteroatoms if tricyclic, said heteroatoms selected from O, N, or S (e.g., carbon atoms and 1-3, 1-6, or 1-9 heteroatoms of N, O, or S if monocyclic, bicyclic, or tricyclic, respectively), wherein any ring atom capable of substitution can be substituted by a substituent. The heterocyclyl groups herein described may also contain fused rings. Fused rings are rings that share a common carbon-carbon bond. Examples of

Attorney's Docket No.: 14174-072W01

heterocyclyl include, but are not limited to tetrahydrofuranyl, tetrahydropyranyl, piperidinyl, morpholino, pyrrolinyl and pyrrolidinyl.

The term "heteroaryl" refers to an aromatic 5-8 membered monocyclic, 8-12 membered bicyclic, or 11-14 membered tricyclic ring system having 1-3 heteroatoms if monocyclic, 1-6
5 heteroatoms if bicyclic, or 1-9 heteroatoms if tricyclic, said heteroatoms selected from O, N, or S (e.g., carbon atoms and 1-3, 1-6, or 1-9 heteroatoms of N, O, or S if monocyclic, bicyclic, or tricyclic, respectively), wherein any ring atom capable of substitution can be substituted by a substituent.

The term "oxo" refers to an oxygen atom, which forms a carbonyl when attached to
10 carbon, an N-oxide when attached to nitrogen, and a sulfoxide or sulfone when attached to sulfur.

The term "acyl" refers to an alkylcarbonyl, cycloalkylcarbonyl, arylcarbonyl, heterocyclcarbonyl, or heteroarylcarbonyl substituent, any of which may be further substituted by substituents.

The term "substituents" refers to a group "substituted" on an alkyl, cycloalkyl, alkenyl,
15 alkynyl, heterocyclyl, heterocycloalkenyl, cycloalkenyl, aryl, or heteroaryl group at any atom of that group. Suitable substituents include, without limitation, alkyl, alkenyl, alkynyl, alkoxy, halo, hydroxy, cyano, nitro, amino, SO_3H , sulfate, phosphate, perfluoroalkyl, perfluoroalkoxy, methylenedioxy, ethylenedioxy, carboxyl, oxo, thioxo, imino (alkyl, aryl, aralkyl), S(O)_n alkyl (where n is 0-2), S(O)_n aryl (where n is 0-2), S(O)_n heteroaryl (where n is 0-2), S(O)_n
20 heterocyclyl (where n is 0-2), amine (mono-, di-, alkyl, cycloalkyl, aralkyl, heteroaralkyl, and combinations thereof), ester (alkyl, aralkyl, heteroaralkyl), amide (mono-, di-, alkyl, aralkyl, heteroaralkyl, and combinations thereof), sulfonamide (mono-, di-, alkyl, aralkyl, heteroaralkyl, and combinations thereof), unsubstituted aryl, unsubstituted heteroaryl, unsubstituted
heterocyclyl, and unsubstituted cycloalkyl. In one aspect, the substituents on a group are
25 independently any one single, or any subset of the aforementioned substituents.

The terms "adeninyl, cytosinyl, guaninyl, thyminyl, and uracilyl" and the like refer to radicals of adenine, cytosine, guanine, thymine, and uracil.

As used herein, an "unusual" nucleobase can include any one of the following:

Attorney's Docket No.: 14174-072W01

2-methyladeninyl,

N6-methyladeninyl,

2-methylthio-N6-methyladeninyl,

N6-isopentenyladeninyl,

5 2-methylthio-N6-isopentenyladeninyl,

N6-(cis-hydroxyisopentenyl)adeninyl,

2-methylthio-N6-(cis-hydroxyisopentenyl) adeninyl,

N6-glycinylicarbamoyladeninyl,

N6-threonylicarbamoyladeninyl,

10 2-methylthio-N6-threonyl carbamoyladeninyl,

N6-methyl-N6-threonylicarbamoyladeninyl,

N6-hydroxynorvalylcarbamoyladeninyl,

2-methylthio-N6-hydroxynorvalyl carbamoyladeninyl,

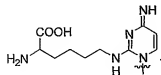
N6,N6-dimethyladeninyl,

15 3-methylcytosinyl,

5-methylcytosinyl,

2-thiocytosinyl,

5-formylcytosinyl,



Attorney's Docket No.: 14174-072W01

N4-methylcytosinyl,

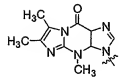
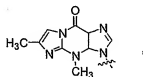
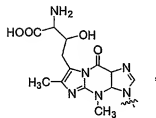
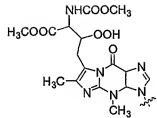
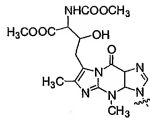
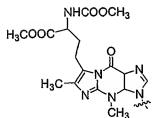
5-hydroxymethylcytosinyl,

1-methylguaninyl,

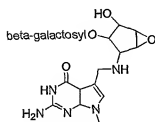
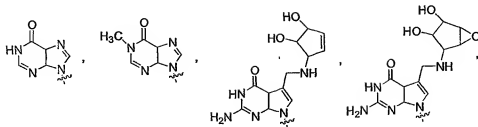
N2-methylguaninyl,

5 7-methylguaninyl,

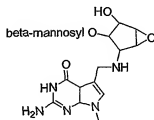
N2,N2-dimethylguaninyl,



Attorney's Docket No.: 14174-072W01



N2,7-dimethylguanylyl,



N2,N2,7-trimethylguanylyl,

1-methylguanylyl,

5 7-cyano-7-deazaguaninylyl,

7-aminomethyl-7-deazaguaninylyl,

pseudouracylyl,

dihydrouracylyl,

5-methyluracylyl,

10 1-methylpseudouracylyl,

2-thiouracylyl,

4-thiouracylyl,

Attorney's Docket No.: 141 74-072 W01

- 2-thiothyminyI
- 5-methyl-2-thiouracilyI,
- 3-(3-amino-3-carboxypropyl)uracilyI,
- 5-hydroxyuracilyI,
- 5 5-methoxyuracilyI,
- uracilyI 5-oxyacetic acid,
- uracilyI 5-oxyacetic acid methyl ester,
- 5-(carboxyhydroxymethyl)uracilyI,
- 5-(carboxyhydroxymethyl)uracilyI methyl ester,
- 10 5-methoxycarbonylmethyluracilyI,
- 5-methoxycarbonylmethyl-2-thiouracilyI,
- 5-aminomethyl-2-thiouracilyI,
- 5-methylaminomethyluracilyI,
- 5-methylaminomethyl-2-thiouracilyI,
- 15 5-methylaminomethyl-2-selenouracilyI,
- 5-carbamoylmethyluracilyI,
- 5-carboxymethylaminomethyluracilyI,
- 5-carboxymethylaminomethyl-2-thiouracilyI,
- 3-methyluracilyI,
- 20 1-methyl-3-(3-amino-3-carboxypropyl) pseudouracilyI,

Attorney's Docket No.: 14174-072W01

5-carboxymethyluracilyl,
5-methyldihydrouracilyl, or
3-methylpseudouracilyl.

Palindromes

5 An RNA, e.g., an iRNA agent, can have a palindrome structure as described herein and those described in one or more of United States Provisional Application Serial No. 60/452,682, filed March 7, 2003; United States Provisional Application Serial No. 60/462,894, filed April 14, 2003; and International Application No. PCT/US04/07070, filed March 8, 2004, all of which are hereby incorporated by reference. The iRNA agents of the invention can target more than
10 one RNA region. For example, an iRNA agent can include a first and second sequence that are sufficiently complementary to each other to hybridize. The first sequence can be complementary to a first target RNA region and the second sequence can be complementary to a second target RNA region. The first and second sequences of the iRNA agent can be on different RNA strands, and the mismatch between the first and second sequences can be less than 50%, 40%,
15 30%, 20%, 10%, 5%, or 1%. The first and second sequences of the iRNA agent are on the same RNA strand, and in a related embodiment more than 50%, 60%, 70%, 80%, 90%, 95%, or 1% of the iRNA agent can be in bimolecular form. The first and second sequences of the iRNA agent can be fully complementary to each other.

The first target RNA region can be encoded by a first gene and the second target RNA
20 region can be encoded by a second gene, or the first and second target RNA regions can be different regions of an RNA from a single gene. The first and second sequences can differ by at least 1 nucleotide.

The first and second target RNA regions can be on transcripts encoded by first and second sequence variants, e.g., first and second alleles, of a gene. The sequence variants can be
25 mutations, or polymorphisms, for example. The first target RNA region can include a nucleotide substitution, insertion, or deletion relative to the second target RNA region, or the second target RNA region can be a mutant or variant of the first target region.

Attorney's Docket No.: 14174-072W01

The first and second target RNA regions can comprise viral or human RNA regions. The first and second target RNA regions can also be on variant transcripts of an oncogene or include different mutations of a tumor suppressor gene transcript. In addition, the first and second target RNA regions can correspond to hot-spots for genetic variation.

5 The compositions of the invention can include mixtures of iRNA agent molecules. For example, one iRNA agent can contain a first sequence and a second sequence sufficiently complementary to each other to hybridize, and in addition the first sequence is complementary to a first target RNA region and the second sequence is complementary to a second target RNA region. The mixture can also include at least one additional iRNA agent variety that includes a
10 third sequence and a fourth sequence sufficiently complementary to each other to hybridize, and where the third sequence is complementary to a third target RNA region and the fourth sequence is complementary to a fourth target RNA region. In addition, the first or second sequence can be sufficiently complementary to the third or fourth sequence to be capable of hybridizing to each other. The first and second sequences can be on the same or different RNA strands, and the third
15 and fourth sequences can be on the same or different RNA strands.

The target RNA regions can be variant sequences of a viral or human RNA, and in certain embodiments, at least two of the target RNA regions can be on variant transcripts of an oncogene or tumor suppressor gene. The target RNA regions can correspond to genetic hot-spots.

20 Methods of making an iRNA agent composition can include obtaining or providing information about a region of an RNA of a target gene (*e.g.*, a viral or human gene, or an oncogene or tumor suppressor, *e.g.*, p53), where the region has high variability or mutational frequency (*e.g.*, in humans). In addition, information about a plurality of RNA targets within the region can be obtained or provided, where each RNA target corresponds to a different variant or
25 mutant of the gene (*e.g.*, a region including the codon encoding p53 248Q and/or p53 249S). The iRNA agent can be constructed such that a first sequence is complementary to a first of the plurality of variant RNA targets (*e.g.*, encoding 249Q) and a second sequence is complementary to a second of the plurality of variant RNA targets (*e.g.*, encoding 249S), and the first and second sequences can be sufficiently complementary to hybridize.

Attorney's Docket No.: 14174-072W01

Sequence analysis, *e.g.*, to identify common mutants in the target gene, can be used to identify a region of the target gene that has high variability or mutational frequency. A region of the target gene having high variability or mutational frequency can be identified by obtaining or providing genotype information about the target gene from a population.

5 Expression of a target gene can be modulated, *e.g.*, downregulated or silenced, by providing an iRNA agent that has a first sequence and a second sequence sufficiently complementary to each other to hybridize. In addition, the first sequence can be complementary to a first target RNA region and the second sequence can be complementary to a second target RNA region.

10 An iRNA agent can include a first sequence complementary to a first variant RNA target region and a second sequence complementary to a second variant RNA target region. The first and second variant RNA target regions can correspond to first and second variants or mutants of a target gene, *e.g.*, viral gene, tumor suppressor or oncogene. The first and second variant target RNA regions can include allelic variants, mutations (*e.g.*, point mutations), or polymorphisms of
15 the target gene. The first and second variant RNA target regions can correspond to genetic hot-spots.

A plurality of iRNA agents (*e.g.*, a panel or bank) can be provided.

Other than Canonical Watson-Crick Duplex Structures

20 An RNA, *e.g.*, an iRNA agent can include monomers which can form other than a canonical Watson-Crick pairing with another monomer, *e.g.*, a monomer on another strand, such as those described herein and those described in United States Provisional Application Serial No. 60/465,665, filed April 25, 2003, and International Application No. PCT/US04/07070, filed March 8, 2004, both of which are hereby incorporated by reference.

25 The use of "other than canonical Watson-Crick pairing" between monomers of a duplex can be used to control, often to promote, melting of all or part of a duplex. The iRNA agent can include a monomer at a selected or constrained position that results in a first level of stability in the iRNA agent duplex (*e.g.*, between the two separate molecules of a double stranded iRNA agent) and a second level of stability in a duplex between a sequence of an iRNA agent and

Attorney's Docket No.: 14174-072W01

another sequence molecule, e.g., a target or off-target sequence in a subject. In some cases the second duplex has a relatively greater level of stability, e.g., in a duplex between an anti-sense sequence of an iRNA agent and a target mRNA. In this case one or more of the monomers, the position of the monomers in the iRNA agent, and the target sequence (sometimes referred to
5 herein as the selection or constraint parameters), are selected such that the iRNA agent duplex is has a comparatively lower free energy of association (which while not wishing to be bound by mechanism or theory, is believed to contribute to efficacy by promoting disassociation of the duplex iRNA agent in the context of the RISC) while the duplex formed between an anti-sense targeting sequence and its target sequence, has a relatively higher free energy of association
10 (which while not wishing to be bound by mechanism or theory, is believed to contribute to efficacy by promoting association of the anti-sense sequence and the target RNA).

In other cases the second duplex has a relatively lower level of stability, e.g., in a duplex between a sense sequence of an iRNA agent and an off-target mRNA. In this case one or more of the monomers, the position of the monomers in the iRNA agent, and an off-target sequence,
15 are selected such that the iRNA agent duplex is has a comparatively higher free energy of association while the duplex formed between a sense targeting sequence and its off-target sequence, has a relatively lower free energy of association (which while not wishing to be bound by mechanism or theory, is believed to reduce the level of off-target silencing by contribute to efficacy by promoting disassociation of the duplex formed by the sense strand and the off-target
20 sequence).

Thus, inherent in the structure of the iRNA agent is the property of having a first stability for the intra-iRNA agent duplex and a second stability for a duplex formed between a sequence from the iRNA agent and another RNA, e.g., a target mRNA. As discussed above, this can be accomplished by judicious selection of one or more of the monomers at a selected or constrained
25 position, the selection of the position in the duplex to place the selected or constrained position, and selection of the sequence of a target sequence (e.g., the particular region of a target gene which is to be targeted). The iRNA agent sequences which satisfy these requirements are sometimes referred herein as constrained sequences. Exercise of the constraint or selection parameters can e, e.g., by inspection, or by computer assisted methods. Exercise of the

Attorney's Docket No.: 14174-072W01

parameters can result in selection of a target sequence and of particular monomers to give a desired result in terms of the stability, or relative stability, of a duplex.

Thus, in another aspect, the invention features, an iRNA agent which includes: a first sequence which targets a first target region and a second sequence which targets a second target
5 region. The first and second sequences have sufficient complementarity to each other to hybridize, e.g., under physiological conditions, e.g., under physiological conditions but not in contact with a helicase or other unwinding enzyme. In a duplex region of the iRNA agent, at a selected or constrained position, the first target region has a first monomer, and the second target region has a second monomer. The first and second monomers occupy complementary or
10 corresponding positions. One, and preferably both monomers are selected such that the stability of the pairing of the monomers contribute to a duplex between the first and second sequence will differ from the stability of the pairing between the first or second sequence with a target sequence.

Usually, the monomers will be selected (selection of the target sequence may be required
15 as well) such that they form a pairing in the iRNA agent duplex which has a lower free energy of dissociation, and a lower T_m , than will be possessed by the pairing of the monomer with its complementary monomer in a duplex between the iRNA agent sequence and a target RNA duplex.

The constraint placed upon the monomers can be applied at a selected site or at more
20 than one selected site. By way of example, the constraint can be applied at more than 1, but less than 3, 4, 5, 6, or 7 sites in an iRNA agent duplex.

A constrained or selected site can be present at a number of positions in the iRNA agent duplex. E.g., a constrained or selected site can be present within 3, 4, 5, or 6 positions from either end, 3' or 5' of a duplexed sequence. A constrained or selected site can be present in the
25 middle of the duplex region, e.g., it can be more than 3, 4, 5, or 6, positions from the end of a duplexed region.

In some embodiment the duplex region of the iRNA agent will have, mismatches, in addition to the selected or constrained site or sites. Preferably it will have no more than 1, 2, 3,

Attorney's Docket No.: 14174-072W01

4, or 5 bases, which do not form canonical Watson-Crick pairs or which do not hybridize. Overhangs are discussed in detail elsewhere herein but are preferably about 2 nucleotides in length. The overhangs can be complementary to the gene sequences being targeted or can be other sequence. TT is a preferred overhang sequence. The first and second iRNA agent
5 sequences can also be joined, e.g., by additional bases to form a hairpin, or by other non-base linkers.

The monomers can be selected such that: first and second monomers are naturally occurring ribonucleotides, or modified ribonucleotides having naturally occurring bases, and when occupying complementary sites either do not pair and have no substantial level of H-
10 bonding, or form a non canonical Watson-Crick pairing and form a non-canonical pattern of H bonding, which usually have a lower free energy of dissociation than seen in a canonical Watson-Crick pairing, or otherwise pair to give a free energy of association which is less than that of a preselected value or is less, e.g., than that of a canonical pairing. When one (or both) of the iRNA agent sequences duplexes with a target, the first (or second) monomer forms a
15 canonical Watson-Crick pairing with the base in the complementary position on the target, or forms a non canonical Watson-Crick pairing having a higher free energy of dissociation and a higher T_m than seen in the pairing in the iRNA agent. The classical Watson-Crick pairings are as follows: A-T, G-C, and A-U. Non-canonical Watson-Crick pairings are known in the art and can include, U-U, G-G, G-A_{trans}, G-A_{cis}, and GU.

20 The monomer in one or both of the sequences is selected such that, it does not pair, or forms a pair with its corresponding monomer in the other sequence which minimizes stability (e.g., the H bonding formed between the monomer at the selected site in the one sequence and its monomer at the corresponding site in the other sequence are less stable than the H bonds formed by the monomer one (or both) of the sequences with the respective target sequence. The
25 monomer is one or both strands is also chosen to promote stability in one or both of the duplexes made by a strand and its target sequence. E.g., one or more of the monomers and the target sequences are selected such that at the selected or constrained position, there is no H bonds formed, or a non canonical pairing is formed in the iRNA agent duplex, or otherwise they otherwise pair to give a free energy of association which is less than that of a preselected value
30 or is less, e.g., than that of a canonical pairing, but when one (or both) sequences form a duplex

Attorney's Docket No.: 14174-072/W01

with the respective target, the pairing at the selected or constrained site is a canonical Watson-Crick pairing.

- 5 The inclusion of such a monomers will have one or more of the following effects: it will destabilize the iRNA agent duplex, it will destabilize interactions between the sense sequence and unintended target sequences, sometimes referred to as off-target sequences, and duplex interactions between the a sequence and the intended target will not be destabilized.

By way of example:

- 10 The monomer at the selected site in the first sequence includes an A (or a modified base which pairs with T), and the monomer in at the selected position in the second sequence is chosen from a monomer which will not pair or which will form a non-canonical pairing, e.g., G. These will be useful in applications wherein the target sequence for the first sequence has a T at the selected position. In embodiments where both target duplexes are stabilized it is useful wherein the target sequence for the second strand has a monomer which will form a canonical Watson-Crick pairing with the monomer selected for the selected position in the second strand.

- 15 The monomer at the selected site in the first sequence includes U (or a modified base which pairs with A), and the monomer in at the selected position in the second sequence is chosen from a monomer which will not pair or which will form a non-canonical pairing, e.g., U or G. These will be useful in applications wherein the target sequence for the first sequence has a T at the selected position. In embodiments where both target duplexes are stabilized it is useful
20 wherein the target sequence for the second strand has a monomer which will form a canonical Watson-Crick pairing with the monomer selected for the selected position in the second strand.

- The monomer at the selected site in the first sequence includes a G (or a modified base which pairs with C), and the monomer in at the selected position in the second sequence is chosen from a monomer which will not pair or which will form a non-canonical pairing, e.g., G,
25 A_{cis}, A_{trans}, or U. These will be useful in applications wherein the target sequence for the first sequence has a T at the selected position. In embodiments where both target duplexes are stabilized it is useful wherein the target sequence for the second strand has a monomer which

Attorney's Docket No.: 14174-072W01

will form a canonical Watson-Crick pairing with the monomer selected for the selected position in the second strand.

The monomer at the selected site in the first sequence includes a C (or a modified base which pairs with G), and the monomer in at the selected position in the second sequence is chosen a monomer which will not pair or which will form a non-canonical pairing. These will be useful in applications wherein the target sequence for the first sequence has a T at the selected position. In embodiments where both target duplexes are stabilized it is useful wherein the target sequence for the second strand has a monomer which will form a canonical Watson-Crick pairing with the monomer selected for the selected position in the second strand.

A non-naturally occurring or modified monomer or monomers can be chosen such that when a non-naturally occurring or modified monomer occupies a positions at the selected or constrained position in an iRNA agent they exhibit a first free energy of dissociation and when one (or both) of them pairs with a naturally occurring monomer, the pair exhibits a second free energy of dissociation, which is usually higher than that of the pairing of the first and second monomers. E.g., when the first and second monomers occupy complementary positions they either do not pair and have no substantial level of H-bonding, or form a weaker bond than one of them would form with a naturally occurring monomer, and reduce the stability of that duplex, but when the duplex dissociates at least one of the strands will form a duplex with a target in which the selected monomer will promote stability, e.g., the monomer will form a more stable pair with a naturally occurring monomer in the target sequence than the pairing it formed in the iRNA agent.

An example of such a pairing is 2-amino A and either of a 2-thio pyrimidine analog of U or T.

When placed in complementary positions of the iRNA agent these monomers will pair very poorly and will minimize stability. However, a duplex is formed between 2 amino A and the U of a naturally occurring target, or a duplex is between 2-thio U and the A of a naturally occurring target or 2-thio T and the A of a naturally occurring target will have a relatively higher free energy of dissociation and be more stable. This is shown in the FIG. 1.

Attorney's Docket No.: 14174-072W01

The pair shown in FIG. 1 (the 2-amino A and the 2-s U and T) is exemplary. In another embodiment, the monomer at the selected position in the sense strand can be a universal pairing moiety. A universal pairing agent will form some level of H bonding with more than one and preferably all other naturally occurring monomers. An examples of a universal pairing moiety is a monomer which includes 3-nitro pyrrole. (Examples of other candidate universal base analogs can be found in the art, e.g., in Loakes, 2001, NAR 29: 2437-2447, hereby incorporated by reference. Examples can also be found in the section on Universal Bases below.) In these cases the monomer at the corresponding position of the anti-sense strand can be chosen for its ability to form a duplex with the target and can include, e.g., A, U, G, or C.

10 iRNA agents of the invention can include:

A sense sequence, which preferably does not target a sequence in a subject, and an anti-sense sequence, which targets a target gene in a subject. The sense and anti-sense sequences have sufficient complementarity to each other to hybridize hybridize, e.g., under physiological conditions, e.g., under physiological conditions but not in contact with a helicase or other
15 unwinding enzyme. In a duplex region of the iRNA agent, at a selected or constrained position, the monomers are selected such that:

The monomer in the sense sequence is selected such that, it does not pair, or forms a pair with its corresponding monomer in the anti-sense strand which minimizes stability (e.g., the H bonding formed between the monomer at the selected site in the sense strand and its monomer at
20 the corresponding site in the anti-sense strand are less stable than the H bonds formed by the monomer of the anti-sense sequence and its canonical Watson-Crick partner or, if the monomer in the anti-sense strand includes a modified base, the natural analog of the modified base and its canonical Watson-Crick partner);

The monomer in the corresponding position in the anti-sense strand is selected such
25 that it maximizes the stability of a duplex it forms with the target sequence, e.g., it forms a canonical Watson-Crick paring with the monomer in the corresponding position on the target stand;

Attorney's Docket No.: 14174-072W01

Optionally, the monomer in the sense sequence is selected such that, it does not pair, or forms a pair with its corresponding monomer in the anti-sense strand which minimizes stability with an off-target sequence.

The inclusion of such a monomers will have one or more of the following effects: it will
5 destabilize the iRNA agent duplex, it will destabilize interactions between the sense sequence and unintended target sequences, sometimes referred to as off-target sequences, and duplex interactions between the anti-sense strand and the intended target will not be destabilized.

The constraint placed upon the monomers can be applied at a selected site or at more than one selected site. By way of example, the constraint can be applied at more than 1, but less
10 than 3, 4, 5, 6, or 7 sites in an iRNA agent duplex.

A constrained or selected site can be present at a number of positions in the iRNA agent duplex. E.g., a constrained or selected site can be present within 3, 4, 5, or 6 positions from either end, 3' or 5' of a duplexed sequence. A constrained or selected site can be present in the middle of the duplex region, e.g., it can be more than 3, 4, 5, or 6, positions from the end of a
15 duplexed region.

In some embodiment the duplex region of the iRNA agent will have, mismatches, in addition to the selected or constrained site or sites. Preferably it will have no more than 1, 2, 3, 4, or 5 bases, which do not form canonical Watson-Crick pairs or which do not hybridize. Overhangs are discussed in detail elsewhere herein but are preferably about 2 nucleotides in
20 length. The overhangs can be complementary to the gene sequences being targeted or can be other sequence. TT is a preferred overhang sequence. The first and second iRNA agent sequences can also be joined, e.g., by additional bases to form a hairpin, or by other non-base linkers.

The monomers can be selected such that: first and second monomers are naturally
25 occurring ribonucleotides, or modified ribonucleotides having naturally occurring bases, and when occupying complementary sites either do not pair and have no substantial level of H-bonding, or form a non canonical Watson-Crick pairing and form a non-canonical pattern of H bonding, which usually have a lower free energy of dissociation than seen in a canonical

Attorney's Docket No.: 14174-072W01

Watson-Crick pairing, or otherwise pair to give a free energy of association which is less than that of a preselected value or is less, e.g., than that of a canonical pairing. When one (or both) of the iRNA agent sequences duplexes with a target, the first (or second) monomer forms a canonical Watson-Crick pairing with the base in the complementary position on the target, or
5 forms a non canonical Watson-Crick pairing having a higher free energy of dissociation and a higher T_m than seen in the pairing in the iRNA agent. The classical Watson-Crick pairings are as follows: A-T, G-C, and A-U. Non-canonical Watson-Crick pairings are known in the art and can include, U-U, G-G, G-A_{trans}, G-A_{cis}, and GU.

The monomer in one or both of the sequences is selected such that, it does not pair, or
10 forms a pair with its corresponding monomer in the other sequence which minimizes stability (e.g., the H bonding formed between the monomer at the selected site in the one sequence and its monomer at the corresponding site in the other sequence are less stable than the H bonds formed by the monomer one (or both) of the sequences with the respective target sequence. The monomer in one or both strands is also chosen to promote stability in one or both of the duplexes
15 made by a strand and its target sequence. E.g., one or more of the monomers and the target sequences are selected such that at the selected or constrained position, there is no H bonds formed, or a non canonical pairing is formed in the iRNA agent duplex, or otherwise they otherwise pair to give a free energy of association which is less than that of a preselected value or is less, e.g., than that of a canonical pairing, but when one (or both) sequences form a duplex
20 with the respective target, the pairing at the selected or constrained site is a canonical Watson-Crick pairing.

The inclusion of such a monomers will have one or more of the following effects: it will destabilize the iRNA agent duplex, it will destabilize interactions between the sense sequence and unintended target sequences, sometimes referred to as off-target sequences, and duplex
25 interactions between the a sequence and the intended target will not be destabilized.

By way of example:

The monomer at the selected site in the first sequence includes an A (or a modified base which pairs with T), and the monomer in at the selected position in the second sequence is chosen from a monomer which will not pair or which will form a non-canonical pairing, e.g., G.

Attorney's Docket No.: 14174-072W01

These will be useful in applications wherein the target sequence for the first sequence has a T at the selected position. In embodiments where both target duplexes are stabilized it is useful wherein the target sequence for the second strand has a monomer which will form a canonical Watson-Crick pairing with the monomer selected for the selected position in the second strand.

- 5 The monomer at the selected site in the first sequence includes U (or a modified base which pairs with A), and the monomer in at the selected position in the second sequence is chosen from a monomer which will not pair or which will form a non-canonical pairing, e.g., U or G. These will be useful in applications wherein the target sequence for the first sequence has a T at the selected position. In embodiments where both target duplexes are stabilized it is useful
10 wherein the target sequence for the second strand has a monomer which will form a canonical Watson-Crick pairing with the monomer selected for the selected position in the second strand.

- The monomer at the selected site in the first sequence includes a G (or a modified base which pairs with C), and the monomer in at the selected position in the second sequence is chosen from a monomer which will not pair or which will form a non-canonical pairing, e.g., G,
15 A_{cis}, A_{trans}, or U. These will be useful in applications wherein the target sequence for the first sequence has a T at the selected position. In embodiments where both target duplexes are stabilized it is useful wherein the target sequence for the second strand has a monomer which will form a canonical Watson-Crick pairing with the monomer selected for the selected position in the second strand.

- 20 The monomer at the selected site in the first sequence includes a C (or a modified base which pairs with G), and the monomer in at the selected position in the second sequence is chosen a monomer which will not pair or which will form a non-canonical pairing. These will be useful in applications wherein the target sequence for the first sequence has a T at the selected position. In embodiments where both target duplexes are stabilized it is useful wherein the target
25 sequence for the second strand has a monomer which will form a canonical Watson-Crick pairing with the monomer selected for the selected position in the second strand.

A non-naturally occurring or modified monomer or monomers can be chosen such that when a non-naturally occurring or modified monomer occupies a positions at the selected or constrained position in an iRNA agent they exhibit a first free energy of dissociation and when

Attorney's Docket No.: 14174-072W01

one (or both) of them pairs with a naturally occurring monomer, the pair exhibits a second free energy of dissociation, which is usually higher than that of the pairing of the first and second monomers. E.g., when the first and second monomers occupy complementary positions they either do not pair and have no substantial level of H-bonding, or form a weaker bond than one of them would form with a naturally occurring monomer, and reduce the stability of that duplex, but when the duplex dissociates at least one of the strands will form a duplex with a target in which the selected monomer will promote stability, e.g., the monomer will form a more stable pair with a naturally occurring monomer in the target sequence than the pairing it formed in the iRNA agent.

An example of such a pairing is 2-amino A and either of a 2-thio pyrimidine analog of U or T.

When placed in complementary positions of the iRNA agent these monomers will pair very poorly and will minimize stability. However, a duplex is formed between 2 amino A and the U of a naturally occurring target, or a duplex is between 2-thio U and the A of a naturally occurring target or 2-thio T and the A of a naturally occurring target will have a relatively higher free energy of dissociation and be more stable.

The monomer at the selected position in the sense strand can be a universal pairing moiety. A universal pairing agent will form some level of H bonding with more than one and preferably all other naturally occurring monomers. An examples of a universal pairing moiety is a monomer which includes 3-nitro pyrrole. (Examples of other candidate universal base analogs can be found in the art, e.g., in Loakes, 2001, NAR 29: 2437-2447, hereby incorporated by reference. Examples can also be found in the section on Universal Bases below.) In these cases the monomer at the corresponding position of the anti-sense strand can be chosen for its ability to form a duplex with the target and can include, e.g., A, U, G, or C.

iRNA agents of the invention can include:

A sense sequence, which preferably does not target a sequence in a subject, and an anti-sense sequence, which targets a target gene in a subject. The sense and anti-sense sequences have sufficient complementarity to each other to hybridize hybridize, e.g., under physiological

Attorney's Docket No.: 14174-072W01

conditions, e.g., under physiological conditions but not in contact with a helicase or other unwinding enzyme. In a duplex region of the iRNA agent, at a selected or constrained position, the monomers are selected such that:

5 The monomer in the sense sequence is selected such that, it does not pair, or forms a pair with its corresponding monomer in the anti-sense strand which minimizes stability (e.g., the H bonding formed between the monomer at the selected site in the sense strand and its monomer at the corresponding site in the anti-sense strand are less stable than the H bonds formed by the monomer of the anti-sense sequence and its canonical Watson-Crick partner or, if the monomer in the anti-sense strand includes a modified base, the natural analog of the modified base and its
10 canonical Watson-Crick partner);

The monomer in the corresponding position in the anti-sense strand is selected such that it maximizes the stability of a duplex it forms with the target sequence, e.g., it forms a canonical Watson-Crick pairing with the monomer in the corresponding position on the target stand;

15 Optionally, the monomer in the sense sequence is selected such that, it does not pair, or forms a pair with its corresponding monomer in the anti-sense strand which minimizes stability with an off-target sequence.

The inclusion of such a monomers will have one or more of the following effects: it will destabilize the iRNA agent duplex, it will destabilize interactions between the sense sequence
20 and unintended target sequences, sometimes referred to as off-target sequences, and duplex interactions between the anti-sense strand and the intended target will not be destabilized.

The constraint placed upon the monomers can be applied at a selected site or at more than one selected site. By way of example, the constraint can be applied at more than 1, but less than 3, 4, 5, 6, or 7 sites in an iRNA agent duplex.

25 A constrained or selected site can be present at a number of positions in the iRNA agent duplex. E.g., a constrained or selected site can be present within 3, 4, 5, or 6 positions from either end, 3' or 5' of a duplexed sequence. A constrained or selected site can be present in the

Attorney's Docket No.: 14174-072W01

middle of the duplex region, e.g., it can be more than 3, 4, 5, or 6, positions from the end of a duplexed region.

The iRNA agent can be selected to target a broad spectrum of genes, including any of the genes described herein.

5 In a preferred embodiment the iRNA agent has an architecture (architecture refers to one or more of overall length, length of a duplex region, the presence, number, location, or length of overhangs, sing strand versus double strand form) described herein.

E.g., the iRNA agent can be less than 30 nucleotides in length, e.g., 21-23 nucleotides. Preferably, the iRNA is 21 nucleotides in length and there is a duplex region of about 19 pairs.
10 In one embodiment, the iRNA is 21 nucleotides in length, and the duplex region of the iRNA is 19 nucleotides. In another embodiment, the iRNA is greater than 30 nucleotides in length.

In some embodiment the duplex region of the iRNA agent will have, mismatches, in addition to the selected or constrained site or sites. Preferably it will have no more than 1, 2, 3, 4, or 5 bases, which do not form canonical Watson-Crick pairs or which do not hybridize.
15 Overhangs are discussed in detail elsewhere herein but are preferably about 2 nucleotides in length. The overhangs can be complementary to the gene sequences being targeted or can be other sequence. TT is a preferred overhang sequence. The first and second iRNA agent sequences can also be joined, e.g., by additional bases to form a hairpin, or by other non-base linkers.

20 One or more selection or constraint parameters can be exercised such that: monomers at the selected site in the sense and anti-sense sequences are both naturally occurring ribonucleotides, or modified ribonucleotides having naturally occurring bases, and when occupying complementary sites in the iRNA agent duplex either do not pair and have no substantial level of H-bonding, or form a non-canonical Watson-Crick pairing and thus form a
25 non-canonical pattern of H bonding, which generally have a lower free energy of dissociation than seen in a Watson-Crick pairing, or otherwise pair to give a free energy of association which is less than that of a preselected value or is less, e.g., than that of a canonical pairing. When one, usually the anti-sense sequence of the iRNA agent sequences forms a duplex with another

Attorney's Docket No.: 14174-072W01

sequence, generally a sequence in the subject, and generally a target sequence, the monomer forms a classic Watson-Crick pairing with the base in the complementary position on the target, or forms a non-canonical Watson-Crick pairing having a higher free energy of dissociation and a higher T_m than seen in the pairing in the iRNA agent. Optionally, when the other sequence of the iRNA agent, usually the sense sequences forms a duplex with another sequence, generally a sequence in the subject, and generally an off-target sequence, the monomer fails to form a canonical Watson-Crick pairing with the base in the complementary position on the off target sequence, e.g., it forms or forms a non-canonical Watson-Crick pairing having a lower free energy of dissociation and a lower T_m .

By way of example:

the monomer at the selected site in the anti-sense stand includes an A (or a modified base which pairs with T), the corresponding monomer in the target is a T, and the sense strand is chosen from a base which will not pair or which will form a noncanonical pair, e.g., G;

the monomer at the selected site in the anti-sense stand includes a U (or a modified base which pairs with A), the corresponding monomer in the target is an A, and the sense strand is chosen from a monomer which will not pair or which will form a non-canonical pairing, e.g., U or G;

the monomer at the selected site in the anti-sense stand includes a C (or a modified base which pairs with G), the corresponding monomer in the target is a G, and the sense strand is chosen a monomer which will not pair or which will form a non-canonical pairing, e.g., G, A_{cis}, A_{trans}, or U;

the monomer at the selected site in the anti-sense stand includes a G (or a modified base which pairs with C), the corresponding monomer in the target is a C, and the sense strand is chosen from a monomer which will not pair or which will form a non-canonical pairing.

In another embodiment a non-naturally occurring or modified monomer or monomers is chosen such that when it occupies complementary a position in an iRNA agent they exhibit a first free energy of dissociation and when one (or both) of them pairs with a naturally occurring monomer, the pair exhibits a second free energy of dissociation, which is usually higher than that

Attorney's Docket No.: 14174-072W01

of the pairing of the first and second monomers. E.g., when the first and second monomers occupy complementary positions they either do not pair and have no substantial level of H-bonding, or form a weaker bond than one of them would form with a naturally occurring monomer, and reduce the stability of that duplex, but when the duplex dissociates at least one of the strands will form a duplex with a target in which the selected monomer will promote stability, e.g., the monomer will form a more stable pair with a naturally occurring monomer in the target sequence than the pairing it formed in the iRNA agent.

An example of such a pairing is 2-amino A and either of a 2-thio pyrimidine analog of U or T. As is discussed above, when placed in complementary positions of the iRNA agent these monomers will pair very poorly and will minimize stability. However, a duplex is formed between 2 amino A and the U of a naturally occurring target, or a duplex is formed between 2-thio U and the A of a naturally occurring target or 2-thio T and the A of a naturally occurring target will have a relatively higher free energy of dissociation and be more stable.

The monomer at the selected position in the sense strand can be a universal pairing moiety. A universal pairing agent will form some level of H bonding with more than one and preferably all other naturally occurring monomers. An examples of a universal pairing moiety is a monomer which includes 3-nitro pyrrole. Examples of other candidate universal base analogs can be found in the art, e.g., in Loakes, 2001, NAR 29: 2437-2447, hereby incorporated by reference. In these cases the monomer at the corresponding position of the anti-sense strand can be chosen for its ability to form a duplex with the target and can include, e.g., A, U, G, or C.

In another aspect, the invention features, an iRNA agent which includes:

a sense sequence, which preferably does not target a sequence in a subject, and an anti-sense sequence, which targets a plurality of target sequences in a subject, wherein the targets differ in sequence at only 1 or a small number, e.g., no more than 5, 4, 3 or 2 positions. The sense and anti-sense sequences have sufficient complementarity to each other to hybridize, e.g., under physiological conditions, e.g., under physiological conditions but not in contact with a helicase or other unwinding enzyme. In the sequence of the anti-sense strand of the iRNA agent is selected such that at one, some, or all of the positions which correspond to positions that differ in sequence between the target sequences, the anti-sense strand will include a monomer

Attorney's Docket No.: 14174-072W01

which will form H-bonds with at least two different target sequences. In a preferred example the anti-sense sequence will include a universal or promiscuous monomer, e.g., a monomer which includes 5-nitro pyrrole, 2-amino A, 2-thio U or 2-thio T, or other universal base referred to herein.

- 5 In a preferred embodiment the iRNA agent targets repeated sequences (which differ at only one or a small number of positions from each other) in a single gene, a plurality of genes, or a viral genome, e.g., the HCV genome.

An embodiment is illustrated in the FIGs. 2 and 3.

- 10 In another aspect, the invention features, determining, e.g., by measurement or calculation, the stability of a pairing between monomers at a selected or constrained position in the iRNA agent duplex, and preferably determining the stability for the corresponding pairing in a duplex between a sequence from the iRNA agent and another RNA, e.g., a target sequence. The determinations can be compared. An iRNA agent thus analysed can be used in the development of a further modified iRNA agent or can be administered to a subject. This analysis can be
15 performed successively to refine or design optimized iRNA agents.

In another aspect, the invention features, a kit which includes one or more of the following an iRNA described herein, a sterile container in which the iRNA agent is disclosed, and instructions for use.

- 20 In another aspect, the invention features, an iRNA agent containing a constrained sequence made by a method described herein. The iRNA agent can target one or more of the genes referred to herein.

- iRNA agents having constrained or selected sites, e.g., as described herein, can be used in any way described herein. Accordingly, they iRNA agents having constrained or selected sites, e.g., as described herein, can be used to silence a target, e.g., in any of the methods
25 described herein and to target any of the genes described herein or to treat any of the disorders described herein. iRNA agents having constrained or selected sites, e.g., as described herein, can be incorporated into any of the formulations or preparations, e.g., pharmaceutical or sterile

Attorney's Docket No.: 14174-072W01

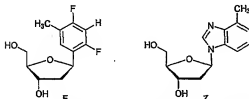
preparations described herein. iRNA agents having constrained or selected sites, e.g., as described herein, can be administered by any of the routes of administration described herein.

The term "other than canonical Watson-Crick pairing" as used herein, refers to a pairing between a first monomer in a first sequence and a second monomer at the corresponding position in a second sequence of a duplex in which one or more of the following is true: (1) there is essentially no pairing between the two, e.g., there is no significant level of H bonding between the monomers or binding between the monomers does not contribute in any significant way to the stability of the duplex; (2) the monomers are a non-canonical pairing of monomers having a naturally occurring bases, i.e., they are other than A-T, A-U, or G-C, and they form monomer-monomer H bonds, although generally the H bonding pattern formed is less strong than the bonds formed by a canonical pairing; or (3) at least one of the monomers includes a non-naturally occurring bases and the H bonds formed between the monomers is, preferably formed is less strong than the bonds formed by a canonical pairing, namely one or more of A-T, A-U, G-C.

The term "off-target" as used herein, refers to as a sequence other than the sequence to be silenced.

Universal Bases: "wild-cards"; shape-based complementarity

Bi-stranded, multisite replication of a base pair between difluorotoluene and adenine: confirmation by 'inverse' sequencing. Liu, D.; Moran, S.; Kool, E. T. *Chem. Biol.*, 1997, 4, 919-926)



(Importance of terminal base pair hydrogen-bonding in 3'-end proofreading by the Klenow fragment of DNA polymerase I. Morales, J. C.; Kool, E. T. *Biochemistry*, 2000, 39, 2626-2632)

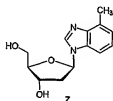
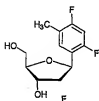
(Selective and stable DNA base pairing without hydrogen bonds. Matray, T, J.; Kool, E. T. *J. Am. Chem. Soc.*, 1998, 120, 6191-6192)

Attorney's Docket No.: 14174-072W01

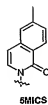
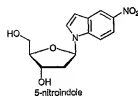
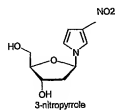


(Difluorotoluene, a nonpolar isostere for thymine, codes specifically and efficiently for adenine in DNA replication. Moran, S. Ren, R. X.-F.; Rumney IV, S.; Kool, E. T. *J. Am. Chem. Soc.*, **1997**, *119*, 2056-2057)

- 5 (Structure and base pairing properties of a replicable nonpolar isostere for deoxyadenosine. Guckian, K. M.; Morales, J. C.; Kool, E. T. *J. Org. Chem.*, **1998**, *63*, 9652-9656)



Attorney's Docket No.: 14174-072/W01



(Universal bases for hybridization, replication and chain termination. Berger, M.; Wu, Y.; Ogawa, A. K.;

- 5 McMinn, D. L.; Schultz, P.G.; Romesberg, F. E. *Nucleic Acids Res.*, **2000**, *28*, 2911-2914)



Attorney's Docket No.: 14174-072W01

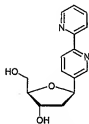
- (1. Efforts toward the expansion of the genetic alphabet: Information storage and replication with unnatural hydrophobic base pairs. Ogawa, A. K.; Wu, Y.; McMin, D. L.; Liu, J.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.*, **2000**, *122*, 3274-3287. 2. Rational design of an unnatural base pair with increased kinetic selectivity. Ogawa, A. K.; Wu, Y.; Berger, M.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.*, **2000**, *122*, 8803-8804)

5



- (Efforts toward expansion of the genetic alphabet: replication of DNA with three base pairs. Tae, E. L.; Wu, Y.; Xia, G.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.*, **2001**, *123*, 7439-7440)

10



- (1. Efforts toward expansion of the genetic alphabet: Optimization of interbase hydrophobic interactions. Wu, Y.; Ogawa, A. K.; Berger, M.; McMin, D. L.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.*, **2000**, *122*, 7621-7632. 2. Efforts toward expansion of genetic alphabet: DNA polymerase recognition of a highly stable, self-pairing hydrophobic base. McMin, D. L.; Ogawa, A. K.; Wu, Y.; Liu, J.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.*, **1999**, *121*, 11585-11586)

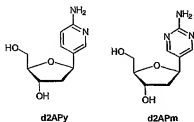
15

(A stable DNA duplex containing a non-hydrogen-bonding and non-shape complementary base couple: Interstrand stacking as the stability determining factor. Brotschi, C.; Haberli, A.; Leumann, C. *J. Angew. Chem. Int. Ed.*, **2001**, *40*, 3012-3014)

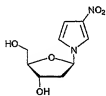
- (2,2'-Bipyridine Ligandose: A novel building block for modifying DNA with intra-duplex metal complexes. Weizman, H.; Tor, Y. *J. Am. Chem. Soc.*, **2001**, *123*, 3375-3376)

20

Attorney's Docket No.: 14174-072W01

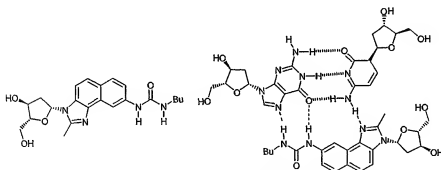


(Minor groove hydration is critical to the stability of DNA duplexes. Lan, T.; McLaughlin, L. W. *J. Am. Chem. Soc.*, **2000**, *122*, 6512-13)



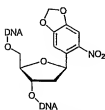
- (Effect of the Universal base 3-nitropyrrole on the selectivity of neighboring natural bases. Oliver, J. S.; Parker, K. A.; Suggs, J. W. *Organic Lett.*, **2001**, *3*, 1977-1980. 2. Effect of the 1-(2'-deoxy-β-D-ribofuranosyl)-3-nitropyrrol residue on the stability of DNA duplexes and triplexes. Amosova, O.; George J.; Fresco, J. R. *Nucleic Acids Res.*, **1997**, *25*, 1930-1934. 3. Synthesis, structure and deoxyribonucleic acid sequencing with a universal nucleosides: 1-(2'-deoxy-β-D-ribofuranosyl)-3-nitropyrrole. Bergstrom, D. E.; Zhang, P.; Toma, P. H.; Andrews, P. C.; Nichols, R. *J. Am. Chem. Soc.*, **1995**, *117*, 1201-1209)

10

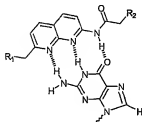


(Model studies directed toward a general triplex DNA recognition scheme: a novel DNA base that binds a CG base-pair in an organic solvent. Zimmerman, S. C.; Schmitt, P. *J. Am. Chem. Soc.*, **1995**, *117*, 10769-10770)

Attorney's Docket No.: 14174-072W01

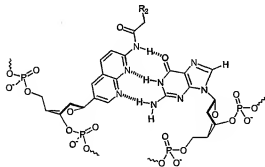
(A universal, photocleavable DNA base: nitropiperonyl 2'-deoxyriboside. *J. Org. Chem.*, 2001, 66, 2067-

2071)



5

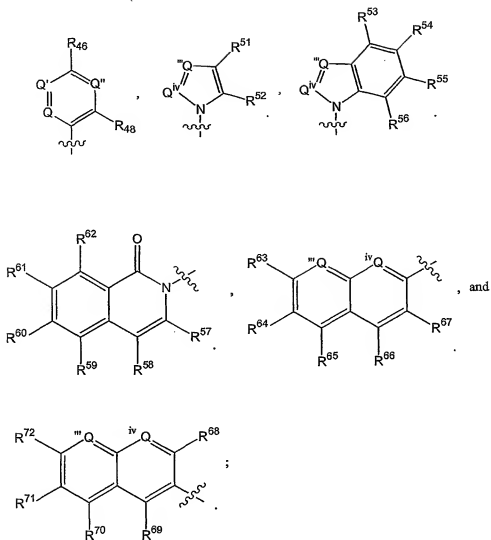
(Recognition of a single guanine bulge by 2-acylamino-1,8-naphthyridine. Nakatani, K.; Sando, S.; Saito, I. *J. Am. Chem. Soc.*, 2000, 122, 2172-2177. b. Specific binding of 2-amino-1,8-naphthyridine into single guanine bulge as evidenced by photooxidation of GC doublet, Nakatani, K.; Sando, S.; Yoshida, K.; Saito, I. *Bioorg. Med. Chem. Lett.*, 2001, 11, 335-337)



10

Other universal bases can have the following formulas:

Attorney's Docket No.: 14174-072W01



wherein:

Q is N or CR⁴⁴;Q' is N or CR⁴⁵;Q'' is N or CR⁴⁷;

Attorney's Docket No.: 14174-072W01

Qⁱⁱⁱ is N or CR⁴⁹;

Q^{iv} is N or CR⁵⁰;

R⁴⁴ is hydrogen, halo, hydroxy, nitro, protected hydroxy, NH₂, NHR^b, or NR^bR^c, C₁-C₆ alkyl, C₆-C₁₀ aryl, C₆-C₁₀ heteroaryl, C₃-C₈ heterocyclyl, or when taken together with R⁴⁵ forms
 5 -OCH₂O-;

R⁴⁵ is hydrogen, halo, hydroxy, nitro, protected hydroxy, NH₂, NHR^b, or NR^bR^c, C₁-C₆ alkyl, C₆-C₁₀ aryl, C₆-C₁₀ heteroaryl, C₃-C₈ heterocyclyl, or when taken together with R⁴⁴ or R⁴⁶ forms -OCH₂O-;

R⁴⁶ is hydrogen, halo, hydroxy, nitro, protected hydroxy, NH₂, NHR^b, or NR^bR^c, C₁-C₆ alkyl, C₆-C₁₀ aryl, C₆-C₁₀ heteroaryl, C₃-C₈ heterocyclyl, or when taken together with R⁴⁵ or R⁴⁷
 10 forms -OCH₂O-;

R⁴⁷ is hydrogen, halo, hydroxy, nitro, protected hydroxy, NH₂, NHR^b, or NR^bR^c, C₁-C₆ alkyl, C₆-C₁₀ aryl, C₆-C₁₀ heteroaryl, C₃-C₈ heterocyclyl, or when taken together with R⁴⁶ or R⁴⁸ forms -OCH₂O-;

R⁴⁸ is hydrogen, halo, hydroxy, nitro, protected hydroxy, NH₂, NHR^b, or NR^bR^c, C₁-C₆ alkyl, C₆-C₁₀ aryl, C₆-C₁₀ heteroaryl, C₃-C₈ heterocyclyl, or when taken together with R⁴⁷ forms
 15 -OCH₂O-;

R⁴⁹, R⁵⁰, R⁵¹, R⁵², R⁵³, R⁵⁴, R⁵⁷, R⁵⁸, R⁵⁹, R⁶⁰, R⁶¹, R⁶², R⁶³, R⁶⁴, R⁶⁵, R⁶⁶, R⁶⁷, R⁶⁸, R⁶⁹, R⁷⁰, R⁷¹, and R⁷² are each independently selected from hydrogen, halo, hydroxy, nitro, protected
 20 hydroxy, NH₂, NHR^b, or NR^bR^c, C₁-C₆ alkyl, C₂-C₆ alkynyl, C₆-C₁₀ aryl, C₆-C₁₀ heteroaryl, C₃-C₈ heterocyclyl, NC(O)R¹⁷, or NC(O)R^o;

R⁵⁵ is hydrogen, halo, hydroxy, nitro, protected hydroxy, NH₂, NHR^b, or NR^bR^c, C₁-C₆ alkyl, C₂-C₆ alkynyl, C₆-C₁₀ aryl, C₆-C₁₀ heteroaryl, C₃-C₈ heterocyclyl, NC(O)R¹⁷, or NC(O)R^o,
 or when taken together with R⁵⁶ forms a fused aromatic ring which may be optionally
 25 substituted;

Attorney's Docket No.: 14174-072W01

R^{56} is hydrogen, halo, hydroxy, nitro, protected hydroxy, NH_2 , NHR^b , or NR^bR^c , C_1 - C_6 alkyl, C_2 - C_6 alkynyl, C_6 - C_{10} aryl, C_6 - C_{10} heteroaryl, C_3 - C_8 heterocyclyl, $NC(O)R^{17}$, or $NC(O)R^o$, or when taken together with R^{55} forms a fused aromatic ring which may be optionally substituted;

5 R^{17} is halo, NH_2 , NHR^b , or NR^bR^c ;

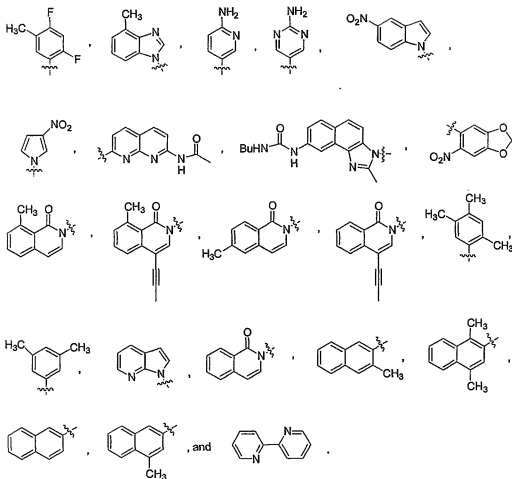
R^b is C_1 - C_6 alkyl or a nitrogen protecting group;

R^c is C_1 - C_6 alkyl; and

R^o is alkyl optionally substituted with halo, hydroxy, nitro, protected hydroxy, NH_2 , NHR^b , or NR^bR^c , C_1 - C_6 alkyl, C_2 - C_6 alkynyl, C_6 - C_{10} aryl, C_6 - C_{10} heteroaryl, C_3 - C_8 heterocyclyl,
10 $NC(O)R^{17}$, or $NC(O)R^o$.

Attorney's Docket No.: 14174-072W01

Examples of universal bases include:



Attorney's Docket No.: 14174-072W01

Asymmetrical Modifications

An RNA, e.g., an iRNA agent, can be asymmetrically modified as described herein, and as described in International Application Serial No. PCT/US04/07070, filed March 8, 2004, which is hereby incorporated by reference.

5 In addition, the invention includes iRNA agents having asymmetrical modifications and another element described herein. E.g., the invention includes an iRNA agent described herein, e.g., a palindromic iRNA agent, an iRNA agent having a non canonical pairing, an iRNA agent which targets a gene described herein, e.g., a gene active in the liver, an iRNA agent having an architecture or structure described herein, an iRNA associated with an amphipathic delivery
10 agent described herein, an iRNA associated with a drug delivery module described herein, an iRNA agent administered as described herein, or an iRNA agent formulated as described herein, which also incorporates an asymmetrical modification.

An asymmetrically modified iRNA agent is one in which a strand has a modification which is not present on the other strand. An asymmetrical modification is a modification found
15 on one strand but not on the other strand. Any modification, e.g., any modification described herein, can be present as an asymmetrical modification. An asymmetrical modification can confer any of the desired properties associated with a modification, e.g., those properties discussed herein. E.g., an asymmetrical modification can: confer resistance to degradation, an alteration in half life; target the iRNA agent to a particular target, e.g., to a particular tissue;
20 modulate, e.g., increase or decrease, the affinity of a strand for its complement or target sequence; or hinder or promote modification of a terminal moiety, e.g., modification by a kinase or other enzymes involved in the RISC mechanism pathway. The designation of a modification as having one property does not mean that it has no other property, e.g., a modification referred to as one which promotes stabilization might also enhance targeting.

25 While not wishing to be bound by theory or any particular mechanistic model, it is believed that asymmetrical modification allows an iRNA agent to be optimized in view of the different or "asymmetrical" functions of the sense and antisense strands. For example, both strands can be modified to increase nuclease resistance, however, since some changes can inhibit RISC activity, these changes can be chosen for the sense strand. In addition, since some

Attorney's Docket No.: 14174-072W01

modifications, e.g., targeting moieties, can add large bulky groups that, e.g., can interfere with the cleavage activity of the RISC complex, such modifications are preferably placed on the sense strand. Thus, targeting moieties, especially bulky ones (e.g. cholesterol), are preferentially added to the sense strand. In one embodiment, an asymmetrical modification in which a phosphate of
5 the backbone is substituted with S, e.g., a phosphorothioate modification, is present in the antisense strand, and a 2' modification, e.g., 2' OMe is present in the sense strand. A targeting moiety can be present at either (or both) the 5' or 3' end of the sense strand of the iRNA agent. In a preferred example, a P of the backbone is replaced with S in the antisense strand, 2'OMe is present in the sense strand, and a targeting moiety is added to either the 5' or 3' end of the sense
10 strand of the iRNA agent.

In a preferred embodiment an asymmetrically modified iRNA agent has a modification on the sense strand which modification is not found on the antisense strand and the antisense strand has a modification which is not found on the sense strand.

Each strand can include one or more asymmetrical modifications. By way of example:
15 one strand can include a first asymmetrical modification which confers a first property on the iRNA agent and the other strand can have a second asymmetrical modification which confers a second property on the iRNA. E.g., one strand, e.g., the sense strand can have a modification which targets the iRNA agent to a tissue, and the other strand, e.g., the antisense strand, has a modification which promotes hybridization with the target gene sequence.

20 In some embodiments both strands can be modified to optimize the same property, e.g., to increase resistance to nucleolytic degradation, but different modifications are chosen for the sense and the antisense strands, e.g., because the modifications affect other properties as well. E.g., since some changes can affect RISC activity these modifications are chosen for the sense strand.

25 In an embodiment one strand has an asymmetrical 2' modification, e.g., a 2' OMe modification, and the other strand has an asymmetrical modification of the phosphate backbone, e.g., a phosphorothioate modification. So, in one embodiment the antisense strand has an asymmetrical 2' OMe modification and the sense strand has an asymmetrical phosphorothioate modification (or vice versa). In a particularly preferred embodiment the RNAi agent will have

Attorney's Docket No.: 14174-072W01

asymmetrical 2'-O alkyl, preferably, 2'-OMe modifications on the sense strand and asymmetrical backbone P modification, preferably a phosphothioate modification in the antisense strand. There can be one or multiple 2'-OMe modifications, e.g., at least 2, 3, 4, 5, or 6, of the subunits of the sense strand can be so modified. There can be one or multiple
5 phosphorothioate modifications, e.g., at least 2, 3, 4, 5, or 6, of the subunits of the antisense strand can be so modified. It is preferable to have an iRNA agent wherein there are multiple 2'-OMe modifications on the sense strand and multiple phosphorothioate modifications on the antisense strand. All of the subunits on one or both strands can be so modified. A particularly preferred embodiment of multiple asymmetric modification on both strands has a duplex region
10 about 20-21, and preferably 19, subunits in length and one or two 3' overhangs of about 2 subunits in length.

Asymmetrical modifications are useful for promoting resistance to degradation by nucleases, e.g., endonucleases. iRNA agents can include one or more asymmetrical modifications which promote resistance to degradation. In preferred embodiments the
15 modification on the antisense strand is one which will not interfere with silencing of the target, e.g., one which will not interfere with cleavage of the target. Most if not all sites on a strand are vulnerable, to some degree, to degradation by endonucleases. One can determine sites which are relatively vulnerable and insert asymmetrical modifications which inhibit degradation. It is often desirable to provide asymmetrical modification of a UA site in an iRNA agent, and in some
20 cases it is desirable to provide the UA sequence on both strands with asymmetrical modification. Examples of modifications which inhibit endonucleolytic degradation can be found herein. Particularly favored modifications include: 2' modification, e.g., provision of a 2' OMe moiety on the U, especially on a sense strand; modification of the backbone, e.g., with the replacement of an O with an S, in the phosphate backbone, e.g., the provision of a phosphorothioate
25 modification, on the U or the A or both, especially on an antisense strand; replacement of the U with a C5 amino linker; replacement of the A with a G (sequence changes are preferred to be located on the sense strand and not the antisense strand); and modification of the at the 2', 6', 7', or 8' position. Preferred embodiments are those in which one or more of these modifications are present on the sense but not the antisense strand, or embodiments where the antisense strand has
30 fewer of such modifications.

Attorney's Docket No.: 14174-072W01

Asymmetrical modification can be used to inhibit degradation by exonucleases.

Asymmetrical modifications can include those in which only one strand is modified as well as those in which both are modified. In preferred embodiments the modification on the antisense strand is one which will not interfere with silencing of the target, e.g., one which will not
5 interfere with cleavage of the target. Some embodiments will have an asymmetrical modification on the sense strand, e.g., in a 3' overhang, e.g., at the 3' terminus, and on the antisense strand, e.g., in a 3' overhang, e.g., at the 3' terminus. If the modifications introduce moieties of different size it is preferable that the larger be on the sense strand. If the modifications introduce moieties of different charge it is preferable that the one with greater
10 charge be on the sense strand.

Examples of modifications which inhibit exonucleolytic degradation can be found herein. Particularly favored modifications include: 2' modification, e.g., provision of a 2' OMe moiety in a 3' overhang, e.g., at the 3' terminus (3' terminus means at the 3' atom of the molecule or at the most 3' moiety, e.g., the most 3' P or 2' position, as indicated by the context); modification
15 of the backbone, e.g., with the replacement of a P with an S, e.g., the provision of a phosphorothioate modification, or the use of a methylated P in a 3' overhang, e.g., at the 3' terminus; combination of a 2' modification, e.g., provision of a 2' OMe moiety and modification of the backbone, e.g., with the replacement of a P with an S, e.g., the provision of a phosphorothioate modification, or the use of a methylated P, in a 3' overhang, e.g., at the 3'
20 terminus; modification with a 3' alkyl; modification with an abasic pyrrolidine in a 3' overhang, e.g., at the 3' terminus; modification with naproxene, ibuprofen, or other moieties which inhibit degradation at the 3' terminus. Preferred embodiments are those in which one or more of these modifications are present on the sense but not the antisense strand, or embodiments where the antisense strand has fewer of such modifications.

Modifications, e.g., those described herein, which affect targeting can be provided as asymmetrical modifications. Targeting modifications which can inhibit silencing, e.g., by inhibiting cleavage of a target, can be provided as asymmetrical modifications of the sense strand. A biodistribution altering moiety, e.g., cholesterol, can be provided in one or more, e.g., two, asymmetrical modifications of the sense strand. Targeting modifications which introduce
30 moieties having a relatively large molecular weight, e.g., a molecular weight of more than 400,

Attorney's Docket No.: 14174-072W01

500, or 1000 daltons, or which introduce a charged moiety (e.g., having more than one positive charge or one negative charge) can be placed on the sense strand.

Modifications, e.g., those described herein, which modulate, e.g., increase or decrease, the affinity of a strand for its complement or target, can be provided as asymmetrical
5 modifications. These include: 5 methyl U; 5 methyl C; pseudouridine, Locked nucleic acids, 2 thio U and 2-amino-A. In some embodiments one or more of these is provided on the antisense strand.

iRNA agents have a defined structure, with a sense strand and an antisense strand, and in many cases short single strand overhangs, e.g., of 2 or 3 nucleotides are present at one or both 3'
10 ends. Asymmetrical modification can be used to optimize the activity of such a structure, e.g., by being placed selectively within the iRNA. E.g., the end region of the iRNA agent defined by the 5' end of the sense strand and the 3' end of the antisense strand is important for function. This region can include the terminal 2, 3, or 4 paired nucleotides and any 3' overhang. In preferred embodiments asymmetrical modifications which result in one or more of the following
15 are used: modifications of the 5' end of the sense strand which inhibit kinase activation of the sense strand, including, e.g., attachments of conjugates which target the molecule or the use of modifications which protect against 5' exonucleolytic degradation; or modifications of either strand, but preferably the sense strand, which enhance binding between the sense and antisense strand and thereby promote a "tight" structure at this end of the molecule.

20 The end region of the iRNA agent defined by the 3' end of the sense strand and the 5' end of the antisense strand is also important for function. This region can include the terminal 2, 3, or 4 paired nucleotides and any 3' overhang. Preferred embodiments include asymmetrical modifications of either strand, but preferably the sense strand, which decrease binding between the sense and antisense strand and thereby promote an "open" structure at this end of the
25 molecule. Such modifications include placing conjugates which target the molecule or modifications which promote nuclease resistance on the sense strand in this region. Modification of the antisense strand which inhibit kinase activation are avoided in preferred embodiments.

Exemplary modifications for asymmetrical placement in the sense strand include the following:

(a) backbone modifications, e.g., modification of a backbone P, including replacement of
5 P with S, or P substituted with alkyl or allyl, e.g., Me, and dithioates (S-P=S); these
modifications can be used to promote nuclease resistance;

(b) 2'-O alkyl, e.g., 2'-OMe, 3'-O alkyl, e.g., 3'-OMe (at terminal and/or internal
positions); these modifications can be used to promote nuclease resistance or to enhance binding
of the sense to the antisense strand, the 3' modifications can be used at the 5' end of the sense
10 strand to avoid sense strand activation by RISC;

(c) 2'-5' linkages (with 2'-H, 2'-OH and 2'-OMe and with P=O or P=S) these
modifications can be used to promote nuclease resistance or to inhibit binding of the sense to the
antisense strand, or can be used at the 5' end of the sense strand to avoid sense strand activation
by RISC;

15 (d) L sugars (e.g., L ribose, L-arabinose with 2'-H, 2'-OH and 2'-OMe); these
modifications can be used to promote nuclease resistance or to inhibit binding of the sense to the
antisense strand, or can be used at the 5' end of the sense strand to avoid sense strand activation
by RISC;

(e) modified sugars (e.g., locked nucleic acids (LNA's), hexose nucleic acids (HNA's)
20 and cyclohexene nucleic acids (CeNA's)); these modifications can be used to promote nuclease
resistance or to inhibit binding of the sense to the antisense strand, or can be used at the 5' end of
the sense strand to avoid sense strand activation by RISC;

(f) nucleobase modifications (e.g., C-5 modified pyrimidines, N-2 modified purines, N-7
modified purines, N-6 modified purines), these modifications can be used to promote nuclease
25 resistance or to enhance binding of the sense to the antisense strand;

(g) cationic groups and Zwitterionic groups (preferably at a terminus), these
modifications can be used to promote nuclease resistance;

Attorney's Docket No.: 14174-072W01

(h) conjugate groups (preferably at terminal positions), e.g., naproxen, biotin, cholesterol, ibuprofen, folic acid, peptides, and carbohydrates; these modifications can be used to promote nuclease resistance or to target the molecule, or can be used at the 5' end of the sense strand to avoid sense strand activation by RISC.

5 Exemplary modifications for asymmetrical placement in the antisense strand include the following:

(a) backbone modifications, e.g., modification of a backbone P, including replacement of P with S, or P substituted with alkyl or allyl, e.g., Me, and dithioates (S-P=S);

(b) 2'-O alkyl, e.g., 2'-OMe, (at terminal positions);

10 (c) 2'-5' linkages (with 2'-H, 2'-OH and 2'-OMe) e.g., terminal at the 3' end); e.g., with P=O or P=S preferably at the 3'-end, these modifications are preferably excluded from the 5' end region as they may interfere with RISC enzyme activity such as kinase activity;

(d) L sugars (e.g., L ribose, L-arabinose with 2'-H, 2'-OH and 2'-OMe); e.g., terminal at the 3' end; e.g., with P=O or P=S preferably at the 3'-end, these modifications are preferably
15 excluded from the 5' end region as they may interfere with kinase activity;

(e) modified sugars (e.g., LNA's, HNA's and CeNA's); these modifications are preferably excluded from the 5' end region as they may contribute to unwanted enhancements of pairing between the sense and antisense strands, it is often preferred to have a "loose" structure in the 5' region, additionally, they may interfere with kinase activity;

20 (f) nucleobase modifications (e.g., C-5 modified pyrimidines, N-2 modified purines, N-7 modified purines, N-6 modified purines);

(g) cationic groups and Zwitterionic groups (preferably at a terminus);

cationic groups and Zwitterionic groups at 2'-position of sugar; 3'-position of the sugar; as nucleobase modifications (e.g., C-5 modified pyrimidines, N-2 modified purines, N-7
25 modified purines, N-6 modified purines);

Attorney's Docket No.: 14174-072W01

conjugate groups (preferably at terminal positions), e.g., naproxen, biotin, cholesterol, ibuprofen, folic acid, peptides, and carbohydrates, but bulky groups or generally groups which inhibit RISC activity should be less preferred.

5 The 5'-OH of the antisense strand should be kept free to promote activity. In some preferred embodiments modifications that promote nuclease resistance should be included at the 3' end, particularly in the 3' overhang.

In another aspect, the invention features a method of optimizing, e.g., stabilizing, an iRNA agent. The method includes selecting a sequence having activity, introducing one or more asymmetric modifications into the sequence, wherein the introduction of the asymmetric
10 modification optimizes a property of the iRNA agent but does not result in a decrease in activity.

The decrease in activity can be less than a preselected level of decrease. In preferred embodiments decrease in activity means a decrease of less than 5, 10, 20, 40, or 50 % activity, as compared with an otherwise similar iRNA lacking the introduced modification. Activity can, e.g., be measured in vivo, or in vitro, with a result in either being sufficient to demonstrate the
15 required maintenance of activity.

The optimized property can be any property described herein and in particular the properties discussed in the section on asymmetrical modifications provided herein. The modification can be any asymmetrical modification, e.g., an asymmetric modification described in the section on asymmetrical modifications described herein. Particularly preferred
20 asymmetric modifications are 2'-O alkyl modifications, e.g., 2'-OMe modifications, particularly in the sense sequence, and modifications of a backbone O, particularly phosphorothioate modifications, in the antisense sequence.

In a preferred embodiment a sense sequence is selected and provided with an asymmetrical modification, while in other embodiments an antisense sequence is selected and
25 provided with an asymmetrical modification. In some embodiments both sense and antisense sequences are selected and each provided with one or more asymmetrical modifications.

Attorney's Docket No.: 14174-072W01

Multiple asymmetric modifications can be introduced into either or both of the sense and antisense sequence. A sequence can have at least 2, 4, 6, 8, or more modifications and all or substantially all of the monomers of a sequence can be modified.

Table 3 shows examples having strand I with a selected modification and strand II with a selected modification.

Table 3. Exemplary strand I- and strand II-modifications

Strand I	Strand II
Nuclease Resistance (e.g., 2'-OMe)	Biodistribution (e.g., P=S)
Biodistribution conjugate (e.g., Lipophile)	Protein Binding Functionality (e.g., Naproxen)
Tissue Distribution Functionality (e.g., Carbohydrates)	Cell Targeting Functionality (e.g., Folate for cancer cells)
Tissue Distribution Functionality (e.g., liver Cell Targeting moieties)	Fusogenic Functionality (e.g., Polyethylene imines)
Cancer Cell Targeting (e.g., RGD peptides and imines)	Fusogenic Functionality (e.g., peptides)
Nuclease Resistance (e.g., 2'-OMe)	Increase in binding Affinity (5-Me-C, 5-Me-U, 2-thio-U, 2-amino-A, G-clamp, LNA)
Tissue Distribution Functionality	RISC activity improving Functionality
Helical conformation changing	Tissue Distribution Functionality

Attorney's Docket No.: 14174-072W01

Functionalities

(P=S; lipophile, carbohydrates)

Z-X-Y Architecture

An RNA, e.g., an iRNA agent, can have a Z-X-Y architecture or structure such as those described herein and those described in copending, co-owned United States Provisional Application Serial No. 60/510,246, filed on October 9, 2003, which is hereby incorporated by reference, copending, co-owned United States Provisional Application Serial No. 60/510,318, filed on October 10, 2003, which is hereby incorporated by reference, and copending, co-owned International Application No. PCT/US04/07070, filed March 8, 2004.

In addition, the invention includes iRNA agents having a Z-X-Y structure and another element described herein. E.g., the invention includes an iRNA agent described herein, e.g., a palindromic iRNA agent, an iRNA agent having a non canonical pairing, an iRNA agent which targets a gene described herein, e.g., a gene active in the liver, an iRNA associated with an amphipathic delivery agent described herein, an iRNA associated with a drug delivery module described herein, an iRNA agent administered as described herein, or an iRNA agent formulated as described herein, which also incorporates a Z-X-Y architecture.

Thus, an iRNA agent can have a first segment, the Z region, a second segment, the X region, and optionally a third region, the Y region:



It may be desirable to modify subunits in one or both of Z and/or Y on one hand and X on the other hand. In some cases they will have the same modification or the same class of modification but it will more often be the case that the modifications made in Z and/or Y will differ from those made in X.

The Z region typically includes a terminus of an iRNA agent. The length of the Z region can vary, but will typically be from 2-14, more preferably 2-10, subunits in length. It typically is single stranded, i.e., it will not base pair with bases of another strand, though it may in some embodiments self associate, e.g., to form a loop structure. Such structures can be formed by the

Attorney's Docket No.: 14174-072W01

end of a strand looping back and forming an intrastrand duplex. E.g., 2, 3, 4, 5 or more intra-strand bases pairs can form, having a looped out or connecting region, typically of 2 or more subunits which do not pair. This can occur at one or both ends of a strand. A typical embodiment of a Z region is a single strand overhang, e.g., an overhang of the length described elsewhere herein. The Z region can thus be or include a 3' or 5' terminal single strand. It can be sense or antisense strand but if it is antisense it is preferred that it is a 3'-overhang. Typical inter-subunit bonds in the Z region include: P=O; P=S; S-P=S; P-NR₂; and P-BR₂. Chiral P=X, where X is S, N, or B) inter-subunit bonds can also be present. (These inter-subunit bonds are discussed in more detail elsewhere herein.) Other preferred Z region subunit modifications (also discussed elsewhere herein) can include: 3'-OR, 3'SR, 2'-OMe, 3'-OMe, and 2'OH modifications and moieties; alpha configuration bases; and 2' arabino modifications.

The X region will in most cases be duplexed, in the case of a single strand iRNA agent, with a corresponding region of the single strand, or in the case of a double stranded iRNA agent, with the corresponding region of the other strand. The length of the X region can vary but will typically be between 10-45 and more preferably between 15 and 35 subunits. Particularly preferred region X's will include 17, 18, 19, 29, 21, 22, 23, 24, or 25 nucleotide pairs, though other suitable lengths are described elsewhere herein and can be used. Typical X region subunits include 2'-OH subunits. In typical embodiments phosphate inter-subunit bonds are preferred while phosphorothioate or non-phosphate bonds are absent. Other modifications preferred in the X region include: modifications to improve binding, e.g., nucleobase modifications; cationic nucleobase modifications; and C-5 modified pyrimidines, e.g., allylamines. Some embodiments have 4 or more consecutive 2'OH subunits. While the use of phosphorothioate is sometimes non preferred they can be used if they connect less than 4 consecutive 2'OH subunits.

The Y region will generally conform to the parameters set out for the Z regions. However, the X and Z regions need not be the same, different types and numbers of modifications can be present, and in fact, one will usually be a 3' overhang and one will usually be a 5' overhang.

In a preferred embodiment the iRNA agent will have a Y and/or Z region each having ribonucleosides in which the 2'-OH is substituted, e.g., with 2'-OMe or other alkyl; and an X

Attorney's Docket No.: 14174-072W01

region that includes at least four consecutive ribonucleoside subunits in which the 2'-OH remains unsubstituted.

The subunit linkages (the linkages between subunits) of an iRNA agent can be modified, e.g., to promote resistance to degradation. Numerous examples of such modifications are disclosed herein, one example of which is the phosphorothioate linkage. These modifications can be provided between the subunits of any of the regions, Y, X, and Z. However, it is preferred that their occurrence is minimized and in particular it is preferred that consecutive modified linkages be avoided.

In a preferred embodiment the iRNA agent will have a Y and Z region each having ribonucleosides in which the 2'-OH is substituted, e.g., with 2'-OMe; and an X region that includes at least four consecutive subunits, e.g., ribonucleoside subunits in which the 2'-OH remains unsubstituted.

As mentioned above, the subunit linkages of an iRNA agent can be modified, e.g., to promote resistance to degradation. These modifications can be provided between the subunits of any of the regions, Y, X, and Z. However, it is preferred that they are minimized and in particular it is preferred that consecutive modified linkages be avoided.

Thus, in a preferred embodiment, not all of the subunit linkages of the iRNA agent are modified and more preferably the maximum number of consecutive subunits linked by other than a phosphodiester bond will be 2, 3, or 4. Particularly preferred iRNA agents will not have four or more consecutive subunits, e.g., 2'-hydroxyl ribonucleoside subunits, in which each subunit is joined by modified linkages – i.e. linkages that have been modified to stabilize them from degradation as compared to the phosphodiester linkages that naturally occur in RNA and DNA.

It is particularly preferred to minimize the occurrence in region X. Thus, in preferred embodiments each of the nucleoside subunit linkages in X will be phosphodiester linkages, or if subunit linkages in region X are modified, such modifications will be minimized. E.g., although the Y and/or Z regions can include inter subunit linkages which have been stabilized against degradation, such modifications will be minimized in the X region, and in particular consecutive modifications will be minimized. Thus, in preferred embodiments the maximum number of

Attorney's Docket No.: 14174-072W01

consecutive subunits linked by other than a phosphodiester bond will be 2, 3, or 4. Particular preferred X regions will not have four or more consecutive subunits, e.g., 2'-hydroxyl ribonucleoside subunits, in which each subunits is joined by modified linkages -- i.e. linkages that have been modified to stabilize them from degradation as compared to the phosphodiester linkages that naturally occur in RNA and DNA.

In a preferred embodiment Y and /or Z will be free of phosphorothioate linkages, though either or both may contain other modifications, e.g., other modifications of the subunit linkages.

In a preferred embodiment region X, or in some cases, the entire iRNA agent, has no more than 3 or no more than 4 subunits having identical 2' moieties.

In a preferred embodiment region X, or in some cases, the entire iRNA agent, has no more than 3 or no more than 4 subunits having identical subunit linkages.

In a preferred embodiment one or more phosphorothioate linkages (or other modifications of the subunit linkage) are present in Y and/or Z, but such modified linkages do not connect two adjacent subunits, e.g., nucleosides, having a 2' modification, e.g., a 2'-O-alkyl moiety. E.g., any adjacent 2'-O-alkyl moieties in the Y and/or Z, are connected by a linkage other than a phosphorothioate linkage.

In a preferred embodiment each of Y and/or Z independently has only one phosphorothioate linkage between adjacent subunits, e.g., nucleosides, having a 2' modification, e.g., 2'-O-alkyl nucleosides. If there is a second set of adjacent subunits, e.g., nucleosides, having a 2' modification, e.g., 2'-O-alkyl nucleosides, in Y and/or Z that second set is connected by a linkage other than a phosphorothioate linkage, e.g., a modified linkage other than a phosphorothioate linkage.

In a preferred embodiment each of Y and/or Z independently has more than one phosphorothioate linkage connecting adjacent pairs of subunits, e.g., nucleosides, having a 2' modification, e.g., 2'-O-alkyl nucleosides, but at least one pair of adjacent subunits, e.g., nucleosides, having a 2' modification, e.g., 2'-O-alkyl nucleosides, are be connected by a linkage other than a phosphorothioate linkage, e.g., a modified linkage other than a phosphorothioate linkage.

Attorney's Docket No.: 14174-072W01

In a preferred embodiment one of the above recited limitation on adjacent subunits in Y and/or Z is combined with a limitation on the subunits in X. E.g., one or more phosphorothioate linkages (or other modifications of the subunit linkage) are present in Y and/or Z, but such modified linkages do not connect two adjacent subunits, e.g., nucleosides, having a 2' modification, e.g., a 2'-O-alkyl moiety. E.g., any adjacent 2'-O-alkyl moieties in the Y and/or Z, are connected by a linkage other than a phosphorothioate linkage. In addition, the X region has no more than 3 or no more than 4 identical subunits, e.g., subunits having identical 2' moieties or the X region has no more than 3 or no more than 4 subunits having identical subunit linkages.

A Y and/or Z region can include at least one, and preferably 2, 3 or 4 of a modification disclosed herein. Such modifications can be chosen, independently, from any modification described herein, e.g., from nuclease resistant subunits, subunits with modified bases, subunits with modified intersubunit linkages, subunits with modified sugars, and subunits linked to another moiety, e.g., a targeting moiety. In a preferred embodiment more than 1 of such subunits can be present but in some embodiments it is preferred that no more than 1, 2, 3, or 4 of such modifications occur, or occur consecutively. In a preferred embodiment the frequency of the modification will differ between Y and/or Z and X, e.g., the modification will be present one of Y and/or Z or X and absent in the other.

An X region can include at least one, and preferably 2, 3 or 4 of a modification disclosed herein. Such modifications can be chosen, independently, from any modification described herein, e.g., from nuclease resistant subunits, subunits with modified bases, subunits with modified intersubunit linkages, subunits with modified sugars, and subunits linked to another moiety, e.g., a targeting moiety. In a preferred embodiment more than 1 of such subunits can be present but in some embodiments it is preferred that no more than 1, 2, 3, or 4 of such modifications occur, or occur consecutively.

An RRMS (described elsewhere herein) can be introduced at one or more points in one or both strands of a double-stranded iRNA agent. An RRMS can be placed in a Y and/or Z region, at or near (within 1, 2, or 3 positions) of the 3' or 5' end of the sense strand or at near (within 2 or 3 positions of) the 3' end of the antisense strand. In some embodiments it is preferred to not have an RRMS at or near (within 1, 2, or 3 positions of) the 5' end of the antisense strand. An

Attorney's Docket No.: 14174-072W01

RRMS can be positioned in the X region, and will preferably be positioned in the sense strand or in an area of the antisense strand not critical for antisense binding to the target.

Differential Modification of Terminal Duplex Stability

In one aspect, the invention features an iRNA agent which can have differential
5 modification of terminal duplex stability (DMTDS).

In addition, the invention includes iRNA agents having DMTDS and another element described herein. E.g., the invention includes an iRNA agent described herein, e.g., a palindromic iRNA agent, an iRNA agent having a non canonical pairing, an iRNA agent which targets a gene described herein, e.g., a gene active in the liver, an iRNA agent having an
10 architecture or structure described herein, an iRNA associated with an amphipathic delivery agent described herein, an iRNA associated with a drug delivery module described herein, an iRNA agent administered as described herein, or an iRNA agent formulated as described herein, which also incorporates DMTDS.

iRNA agents can be optimized by increasing the propensity of the duplex to disassociate
15 or melt (decreasing the free energy of duplex association), in the region of the 5' end of the antisense strand duplex. This can be accomplished, e.g., by the inclusion of subunits which increase the propensity of the duplex to disassociate or melt in the region of the 5' end of the antisense strand. It can also be accomplished by the attachment of a ligand that increases the propensity of the duplex to disassociate or melt in the region of the 5' end. While not wishing to
20 be bound by theory, the effect may be due to promoting the effect of an enzyme such as helicase, for example, promoting the effect of the enzyme in the proximity of the 5' end of the antisense strand.

The inventors have also discovered that iRNA agents can be optimized by decreasing the propensity of the duplex to disassociate or melt (increasing the free energy of duplex
25 association), in the region of the 3' end of the antisense strand duplex. This can be accomplished, e.g., by the inclusion of subunits which decrease the propensity of the duplex to disassociate or melt in the region of the 3' end of the antisense strand. It can also be

Attorney's Docket No.: 14174-072/W01

accomplished by the attachment of ligand that decreases the propensity of the duplex to disassociate or melt in the region of the 5' end.

- Modifications which increase the tendency of the 5' end of the duplex to dissociate can be used alone or in combination with other modifications described herein, e.g., with
- 5 modifications which decrease the tendency of the 3' end of the duplex to dissociate. Likewise, modifications which decrease the tendency of the 3' end of the duplex to dissociate can be used alone or in combination with other modifications described herein, e.g., with modifications which increase the tendency of the 5' end of the duplex to dissociate.

Decreasing the stability of the AS 5' end of the duplex

- 10 Subunit pairs can be ranked on the basis of their propensity to promote dissociation or melting (e.g., on the free energy of association or dissociation of a particular pairing, the simplest approach is to examine the pairs on an individual pair basis, though next neighbor or similar analysis can also be used). In terms of promoting dissociation:

A:U is preferred over G:C;

- 15 G:U is preferred over G:C;

I:C is preferred over G:C (I=inosine);

mismatches, e.g., non-canonical or other than canonical pairings (as described elsewhere herein) are preferred over canonical (A:T, A:U, G:C) pairings;

pairings which include a universal base are preferred over canonical pairings.

- 20 A typical ds iRNA agent can be diagrammed as follows:

S 5' R₁ N₁ N₂ N₃ N₄ N₅ [N] N₅ N₄ N₃ N₂ N₁ R₂ 3'

AS 3' R₃ N₁ N₂ N₃ N₄ N₅ [N] N₅ N₄ N₃ N₂ N₁ R₄ 5'

S:AS P₁ P₂ P₃ P₄ P₅ [N] P₅ P₄ P₃ P₂ P₁ 5'

Attorney's Docket No.: 14174-072W01

S indicates the sense strand; AS indicates antisense strand; R₁ indicates an optional (and nonpreferred) 5' sense strand overhang; R₂ indicates an optional (though preferred) 3' sense overhang; R₃ indicates an optional (though preferred) 3' antisense sense overhang; R₄ indicates an optional (and nonpreferred) 5' antisense overhang; N indicates subunits; [N] indicates that
 5 additional subunit pairs may be present; and P_x indicates a pairing of sense N_x and antisense N_x. Overhangs are not shown in the P diagram. In some embodiments a 3' AS overhang corresponds to region Z, the duplex region corresponds to region X, and the 3' S strand overhang corresponds to region Y, as described elsewhere herein. (The diagram is not meant to imply maximum or minimum lengths, on which guidance is provided elsewhere herein.)

10 It is preferred that pairings which decrease the propensity to form a duplex are used at 1 or more of the positions in the duplex at the 5' end of the AS strand. The terminal pair (the most 5' pair in terms of the AS strand) is designated as P₋₁, and the subsequent pairing positions (going in the 3' direction in terms of the AS strand) in the duplex are designated, P₋₂, P₋₃, P₋₄, P₋₅, and so on. The preferred region in which to modify to modulate duplex formation is at P₋₅
 15 through P₋₁, more preferably P₋₄ through P₋₁, more preferably P₋₃ through P₋₁. Modification at P₋₁, is particularly preferred, alone or with modification(s) other position(s), e.g., any of the positions just identified. It is preferred that at least 1, and more preferably 2, 3, 4, or 5 of the pairs of one of the recited regions be chosen independently from the group of:

A:U

20 G:U

I:C

mismatched pairs, e.g., non-canonical or other than canonical pairings or pairings which include a universal base.

In preferred embodiments the change in subunit needed to achieve a pairing which
 25 promotes dissociation will be made in the sense strand, though in some embodiments the change will be made in the antisense strand.

Attorney's Docket No.: 14174-072W01

In a preferred embodiment the at least 2, or 3, of the pairs in P₁, through P₄, are pairs which promote dissociation.

In a preferred embodiment the at least 2, or 3, of the pairs in P₁, through P₄, are A:U.

In a preferred embodiment the at least 2, or 3, of the pairs in P₁, through P₄, are G:U.

5 In a preferred embodiment the at least 2, or 3, of the pairs in P₁, through P₄, are I:C.

In a preferred embodiment the at least 2, or 3, of the pairs in P₁, through P₄, are mismatched pairs, e.g., non-canonical or other than canonical pairings pairings.

In a preferred embodiment the at least 2, or 3, of the pairs in P₁, through P₄, are pairings which include a universal base.

10 *Increasing the stability of the AS 3' end of the duplex*

Subunit pairs can be ranked on the basis of their propensity to promote stability and inhibit dissociation or melting (e.g., on the free energy of association or dissociation of a particular pairing, the simplest approach is to examine the pairs on an individual pair basis, though next neighbor or similar analysis can also be used). In terms of promoting duplex

15 stability:

G:C is preferred over A:U

Watson-Crick matches (A:T, A:U, G:C) are preferred over non-canonical or other than canonical pairings

analogs that increase stability are preferred over Watson-Crick matches (A:T, A:U,

20 G:C)

2-amino-A:U is preferred over A:U

2-thio U or 5 Me-thio-U:A are preferred over U:A

G-clamp (an analog of C having 4 hydrogen bonds):G is preferred over C:G

Attorney's Docket No.: 14174-072W01

guanadinium-G-clamp:G is preferred over C:G

psuedo uridine:A is preferred over U:A

sugar modifications, e.g., 2' modifications, e.g., 2'F, ENA, or LNA, which enhance binding are preferred over non-modified moieties and can be present on one or both strands to enhance stability of the duplex. It is preferred that pairings which increase the propensity to form a duplex are used at 1 or more of the positions in the duplex at the 3' end of the AS strand. The terminal pair (the most 3' pair in terms of the AS strand) is designated as P₁, and the subsequent pairing positions (going in the 5' direction in terms of the AS strand) in the duplex are designated, P₂, P₃, P₄, P₅, and so on. The preferred region in which to modify to modulate duplex formation is at P₃ through P₁, more preferably P₄ through P₁, more preferably P₃ through P₁. Modification at P₁, is particularly preferred, alone or with mdification(s) at other position(s), e.g., any of the positions just identified. It is preferred that at least 1, and more preferably 2, 3, 4, or 5 of the pairs of the recited regions be chosen independently from the group of:

G:C

a pair having an analog that increases stability over Watson-Crick matches (A:T, A:U, G:C)

2-amino-A:U

2-thio U or 5 Me-thio-U:A

G-clamp (an analog of C having 4 hydrogen bonds):G

guanadinium-G-clamp:G

psuedo uridine:A

a pair in which one or both subunits has a sugar modification, e.g., a 2' modification, e.g., 2'F, ENA, or LNA, which enhance binding.

Attorney's Docket No.: 14174-072W01

In a preferred embodiment the at least 2, or 3, of the pairs in P₁, through P₄, are pairs which promote duplex stability.

In a preferred embodiment the at least 2, or 3, of the pairs in P₁, through P₄, are G:C.

In a preferred embodiment the at least 2, or 3, of the pairs in P₁, through P₄, are a pair
5 having an analog that increases stability over Watson-Crick matches.

In a preferred embodiment the at least 2, or 3, of the pairs in P₁, through P₄, are 2-amino-A:U.

In a preferred embodiment the at least 2, or 3, of the pairs in P₁, through P₄, are 2-thio U or 5 Me-thio-U:A.

10 In a preferred embodiment the at least 2, or 3, of the pairs in P₁, through P₄, are G-clamp:G.

In a preferred embodiment the at least 2, or 3, of the pairs in P₁, through P₄, are guanidinium-G-clamp:G.

15 In a preferred embodiment the at least 2, or 3, of the pairs in P₁, through P₄, are psuedo uridine:A.

In a preferred embodiment the at least 2, or 3, of the pairs in P₁, through P₄, are a pair in which one or both subunits has a sugar modification, e.g., a 2' modification, e.g., 2'F, ENA, or LNA, which enhances binding.

G-clamps and guanidinium G-clamps are discussed in the following references: Holmes
20 and Gait, "The Synthesis of 2'-O-Methyl G-Clamp Containing Oligonucleotides and Their Inhibition of the HIV-1 Tat-TAR Interaction," *Nucleosides, Nucleotides & Nucleic Acids*, 22:1259-1262, 2003; Holmes *et al.*, "Steric inhibition of human immunodeficiency virus type-1 Tat-dependent trans-activation in vitro and in cells by oligonucleotides containing 2'-O-methyl G-clamp ribonucleoside analogues," *Nucleic Acids Research*, 31:2759-2768, 2003; Wilds, *et al.*,
25 "Structural basis for recognition of guanosine by a synthetic tricyclic cytosine analogue: Guanidinium G-clamp," *Helvetica Chimica Acta*, 86:966-978, 2003; Rajeev, *et al.*, "High-

Attorney's Docket No.: 14174-072W01

Affinity Peptide Nucleic Acid Oligomers Containing Tricyclic Cytosine Analogues," Organic Letters, 4:4395-4398, 2002; Ausin, *et al.*, "Synthesis of Amino- and Guanidino-G-Clamp PNA Monomers," Organic Letters, 4:4073-4075, 2002; Maier *et al.*, "Nuclease resistance of oligonucleotides containing the tricyclic cytosine analogues phenoxazine and 9-(2-aminoethoxy)-phenoxazine ("G-clamp") and origins of their nuclease resistance properties," 5 Biochemistry, 41:1323-7, 2002; Flanagan, *et al.*, "A cytosine analog that confers enhanced potency to antisense oligonucleotides," Proceedings Of The National Academy Of Sciences Of The United States Of America, 96:3513-8, 1999.

Attorney's Docket No.: 14174-072W01

Simultaneously decreasing the stability of the AS 5' end of the duplex and increasing the stability of the AS 3' end of the duplex

As is discussed above, an iRNA agent can be modified to both decrease the stability of the AS 5' end of the duplex and increase the stability of the AS 3' end of the duplex. This can be effected by combining one or more of the stability decreasing modifications in the AS 5' end of the duplex with one or more of the stability increasing modifications in the AS 3' end of the duplex. Accordingly a preferred embodiment includes modification in P₅ through P₁, more preferably P₄ through P₁ and more preferably P₃ through P₁. Modification at P₁, is particularly preferred, alone or with other position, e.g., the positions just identified. It is preferred that at least 1, and more preferably 2, 3, 4, or 5 of the pairs of one of the recited regions of the AS 5' end of the duplex region be chosen independently from the group of:

A:U

G:U

I:C

15 mismatched pairs, e.g., non-canonical or other than canonical pairings which include a universal base; and

a modification in P₅ through P₁, more preferably P₄ through P₁ and more preferably P₃ through P₁. Modification at P₁, is particularly preferred, alone or with other position, e.g., the positions just identified. It is preferred that at least 1, and more preferably 2, 3, 4, or 5 of the pairs of one of the recited regions of the AS 3' end of the duplex region be chosen independently from the group of:

G:C

a pair having an analog that increases stability over Watson-Crick matches (A:T, A:U, G:C)

25 2-amino-A:U

Attorney's Docket No.: 14174-072W01

2-thio U or 5 Me-thio-U:A

G-clamp (an analog of C having 4 hydrogen bonds):G

guanadinium-G-clamp:G

psuedo uridine:A

- 5 a pair in which one or both subunits has a sugar modification, e.g., a 2' modification, e.g., 2'F, ENA, or LNA, which enhance binding.

The invention also includes methods of selecting and making iRNA agents having DMTDS. E.g., when screening a target sequence for candidate sequences for use as iRNA agents one can select sequences having a DMTDS property described herein or one which can be
10 modified, preferably with as few changes as possible, especially to the

AS strand, to provide a desired level of DMTDS.

The invention also includes, providing a candidate iRNA agent sequence, and modifying at least one P in P₅ through P₁ and/or at least one P in P₅ through P₁ to provide a DMTDS iRNA agent.

- 15 DMTDS iRNA agents can be used in any method described herein, e.g., to silence any gene disclosed herein, to treat any disorder described herein, in any formulation described herein, and generally in and/or with the methods and compositions described elsewhere herein. DMTDS iRNA agents can incorporate other modifications described herein, e.g., the attachment of targeting agents or the inclusion of modifications which enhance stability, e.g., the inclusion of
20 nuclease resistant monomers or the inclusion of single strand overhangs (e.g., 3' AS overhangs and/or 3' S strand overhangs) which self associate to form intrastrand duplex structure.

Preferably these iRNA agents will have an architecture described herein.

Other Embodiments

- An RNA, e.g., an iRNA agent, can be produced in a cell *in vivo*, e.g., from exogenous
25 DNA templates that are delivered into the cell. For example, the DNA templates can be inserted

Attorney's Docket No.: 14174-072W01

into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (U.S. Pat. No. 5,328,470), or by stereotactic injection (see, *e.g.*, Chen *et al.*, *Proc. Natl. Acad. Sci. USA* 91:3054-3057, 1994). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. The DNA templates, for example, can include two transcription units, one that produces a transcript that includes the top strand of an iRNA agent and one that produces a transcript that includes the bottom strand of an iRNA agent. When the templates are transcribed, the iRNA agent is produced, and processed into sRNA agent fragments that mediate gene silencing.

In vivo Delivery

An iRNA agent can be linked, *e.g.*, noncovalently linked to a polymer for the efficient delivery of the iRNA agent to a subject, *e.g.*, a mammal, such as a human. The iRNA agent can, for example, be complexed with cyclodextrin. Cyclodextrins have been used as delivery vehicles of therapeutic compounds. Cyclodextrins can form inclusion complexes with drugs that are able to fit into the hydrophobic cavity of the cyclodextrin. In other examples, cyclodextrins form non-covalent associations with other biologically active molecules such as oligonucleotides and derivatives thereof. The use of cyclodextrins creates a water-soluble drug delivery complex, that can be modified with targeting or other functional groups. Cyclodextrin cellular delivery system for oligonucleotides described in U.S. Pat. No. 5,691,316, which is hereby incorporated by reference, are suitable for use in methods of the invention. In this system, an oligonucleotide is noncovalently complexed with a cyclodextrin, or the oligonucleotide is covalently bound to adamantane which in turn is non-covalently associated with a cyclodextrin.

The delivery molecule can include a linear cyclodextrin copolymer or a linear oxidized cyclodextrin copolymer having at least one ligand bound to the cyclodextrin copolymer. Delivery systems, as described in U.S. Patent No. 6,509,323, herein incorporated by reference, are suitable for use in methods of the invention. An iRNA agent can be bound to the linear cyclodextrin copolymer and/or a linear oxidized cyclodextrin copolymer. Either or both of the

Attorney's Docket No.: 14174-072W01

cyclodextrin or oxidized cyclodextrin copolymers can be crosslinked to another polymer and/or bound to a ligand.

A composition for iRNA delivery can employ an "inclusion complex," a molecular compound having the characteristic structure of an adduct. In this structure, the "host

5 molecule" spatially encloses at least part of another compound in the delivery vehicle. The enclosed compound (the "guest molecule") is situated in the cavity of the host molecule without affecting the framework structure of the host. A "host" is preferably cyclodextrin, but can be any of the molecules suggested in U.S. Patent Publ. 2003/0008818, herein incorporated by reference.

Cyclodextrins can interact with a variety of ionic and molecular species, and the resulting
10 inclusion compounds belong to the class of "host-guest" complexes. Within the host-guest relationship, the binding sites of the host and guest molecules should be complementary in the stereoelectronic sense. A composition of the invention can contain at least one polymer and at least one therapeutic agent, generally in the form of a particulate composite of the polymer and therapeutic agent, e.g., the iRNA agent. The iRNA agent can contain one or more complexing
15 agents. At least one polymer of the particulate composite can interact with the complexing agent in a host-guest or a guest-host interaction to form an inclusion complex between the polymer and the complexing agent. The polymer and, more particularly, the complexing agent can be used to introduce functionality into the composition. For example, at least one polymer of the particulate composite has host functionality and forms an inclusion complex with a complexing agent
20 having guest functionality. Alternatively, at least one polymer of the particulate composite has guest functionality and forms an inclusion complex with a complexing agent having host functionality. A polymer of the particulate composite can also contain both host and guest functionalities and form inclusion complexes with guest complexing agents and host complexing agents. A polymer with functionality can, for example, facilitate cell targeting and/or cell
25 contact (e.g., targeting or contact to a liver cell), intercellular trafficking, and/or cell entry and release.

Upon forming the particulate composite, the iRNA agent may or may not retain its biological or therapeutic activity. Upon release from the therapeutic composition, specifically, from the polymer of the particulate composite, the activity of the iRNA agent is restored.

Attorney's Docket No.: 14174-072W01

Accordingly, the particulate composite advantageously affords the iRNA agent protection against loss of activity due to, for example, degradation and offers enhanced bioavailability. Thus, a composition may be used to provide stability, particularly storage or solution stability, to an iRNA agent or any active chemical compound. The iRNA agent may be further modified
5 with a ligand prior to or after particulate composite or therapeutic composition formation. The ligand can provide further functionality. For example, the ligand can be a targeting moiety.

Physiological Effects

The iRNA agents described herein can be designed such that determining therapeutic toxicity is made easier by the complementarity of the iRNA agent with both a human and a non-
10 human animal sequence. By these methods, an iRNA agent can consist of a sequence that is fully complementary to a nucleic acid sequence from a human and a nucleic acid sequence from at least one non-human animal, *e.g.*, a non-human mammal, such as a rodent, ruminant or primate. For example, the non-human mammal can be a mouse, rat, dog, pig, goat, sheep, cow, monkey, Pan paniscus, Pan troglodytes, Macaca mulatto, or Cynomolgus monkey. The sequence
15 of the iRNA agent could be complementary to sequences within homologous genes, *e.g.*, oncogenes or tumor suppressor genes, of the non-human mammal and the human. By determining the toxicity of the iRNA agent in the non-human mammal, one can extrapolate the toxicity of the iRNA agent in a human. For a more strenuous toxicity test, the iRNA agent can be complementary to a human and more than one, *e.g.*, two or three or more, non-human
20 animals.

The methods described herein can be used to correlate any physiological effect of an iRNA agent on a human, *e.g.*, any unwanted effect, such as a toxic effect, or any positive, or desired effect.

Delivery Module

25 An RNA, *e.g.*, an iRNA agent described herein, can be used with a drug delivery conjugate or module, such as those described herein and those described in copending, co-owned United States Provisional Application Serial No. 60/454,265, filed on March 12, 2003, and

Attorney's Docket No.: 14174-072W01

International Application Serial No. PCT/US04/07070, filed March 8, 2004, both of which are hereby incorporated by reference.

In addition, the invention includes iRNA agents described herein, e.g., a palindromic iRNA agent, an iRNA agent having a non canonical pairing, an iRNA agent which targets a gene described herein, e.g., a gene active in the liver, an iRNA agent having a chemical modification described herein, e.g., a modification which enhances resistance to degradation, an iRNA agent having an architecture or structure described herein, an iRNA agent administered as described herein, or an iRNA agent formulated as described herein, combined with, associated with, and delivered by such a drug delivery conjugate or module.

The iRNA agents can be complexed to a delivery agent that features a modular complex. The complex can include a carrier agent linked to one or more of (preferably two or more, more preferably all three of): (a) a condensing agent (e.g., an agent capable of attracting, e.g., binding, a nucleic acid, e.g., through ionic or electrostatic interactions); (b) a fusogenic agent (e.g., an agent capable of fusing and/or being transported through a cell membrane, e.g., an endosome membrane); and (c) a targeting group, e.g., a cell or tissue targeting agent, e.g., a lectin, glycoprotein, lipid or protein, e.g., an antibody, that binds to a specified cell type such as a liver cell.

An iRNA agent, e.g., iRNA agent or sRNA agent described herein, can be linked, e.g., coupled or bound, to the modular complex. The iRNA agent can interact with the condensing agent of the complex, and the complex can be used to deliver an iRNA agent to a cell, e.g., *in vitro* or *in vivo*. For example, the complex can be used to deliver an iRNA agent to a subject in need thereof, e.g., to deliver an iRNA agent to a subject having a disorder, e.g., a disorder described herein, such as a disease or disorder of the liver.

The fusogenic agent and the condensing agent can be different agents or the one and the same agent. For example, a polyamino chain, e.g., polyethylencimine (PEI), can be the fusogenic and/or the condensing agent.

The delivery agent can be a modular complex. For example, the complex can include a carrier agent linked to one or more of (preferably two or more, more preferably all three of):

Attorney's Docket No.: 14174-072W01

(a) a condensing agent (*e.g.*, an agent capable of attracting, *e.g.*, binding, a nucleic acid, *e.g.*, through ionic interaction),

(b) a fusogenic agent (*e.g.*, an agent capable of fusing and/or being transported through a cell membrane, *e.g.*, an endosome membrane), and

- 5 (c) a targeting group, *e.g.*, a cell or tissue targeting agent, *e.g.*, a lectin, glycoprotein, lipid or protein, *e.g.*, an antibody, that binds to a specified cell type such as a liver cell. A targeting group can be a thyrotropin, melanotropin, lectin, glycoprotein, surfactant protein A, Mucin carbohydrate, multivalent lactose, multivalent galactose, N-acetyl-galactosamine, N-acetyl-gulcosamine multivalent mannose, multivalent fucose, glycosylated polyaminoacids,
- 10 multivalent galactose, transferrin, bisphosphonate, polyglutamate, polyaspartate, a lipid, cholesterol, a steroid, bile acid, folate, vitamin B12, biotin, Neproxin, or an RGD peptide or RGD peptide mimetic.

Carrier agents

- The carrier agent of a modular complex described herein can be a substrate for
- 15 attachment of one or more of: a condensing agent, a fusogenic agent, and a targeting group. The carrier agent would preferably lack an endogenous enzymatic activity. The agent would preferably be a biological molecule, preferably a macromolecule. Polymeric biological carriers are preferred. It would also be preferred that the carrier molecule be biodegradable..

- The carrier agent can be a naturally occurring substance, such as a protein (*e.g.*, human
- 20 serum albumin (HSA), low-density lipoprotein (LDL), or globulin); carbohydrate (*e.g.*, a dextran, pullulan, chitin, chitosan, inulin, cyclodextrin or hyaluronic acid); or lipid. The carrier molecule can also be a recombinant or synthetic molecule, such as a synthetic polymer, *e.g.*, a synthetic polyamino acid. Examples of polyamino acids include polylysine (PLL), poly L-aspartic acid, poly L-glutamic acid, styrene-maleic acid anhydride copolymer, poly(L-lactide-co-glycolide) copolymer, divinyl ether-maleic anhydride copolymer, N-(2-
- 25 hydroxypropyl)methacrylamide copolymer (HMPA), polyethylene glycol (PEG), polyvinyl alcohol (PVA), polyurethane, poly(2-ethylacrylic acid), N-isopropylacrylamide polymers, or polyphosphazine. Other useful carrier molecules can be identified by routine methods.

Attorney's Docket No.: 14174-072W01

A carrier agent can be characterized by one or more of: (a) is at least 1 Da in size; (b) has at least 5 charged groups, preferably between 5 and 5000 charged groups; (c) is present in the complex at a ratio of at least 1:1 carrier agent to fusogenic agent; (d) is present in the complex at a ratio of at least 1:1 carrier agent to condensing agent; (e) is present in the complex at a ratio of at least 1:1 carrier agent to targeting agent.

Fusogenic agents

A fusogenic agent of a modular complex described herein can be an agent that is responsive to, *e.g.*, changes charge depending on, the pH environment. Upon encountering the pH of an endosome, it can cause a physical change, *e.g.*, a change in osmotic properties which disrupts or increases the permeability of the endosome membrane. Preferably, the fusogenic agent changes charge, *e.g.*, becomes protonated, at pH lower than physiological range. For example, the fusogenic agent can become protonated at pH 4.5-6.5. The fusogenic agent can serve to release the iRNA agent into the cytoplasm of a cell after the complex is taken up, *e.g.*, via endocytosis, by the cell, thereby increasing the cellular concentration of the iRNA agent in the cell.

In one embodiment, the fusogenic agent can have a moiety, *e.g.*, an amino group, which, when exposed to a specified pH range, will undergo a change, *e.g.*, in charge, *e.g.*, protonation. The change in charge of the fusogenic agent can trigger a change, *e.g.*, an osmotic change, in a vesicle, *e.g.*, an endocytic vesicle, *e.g.*, an endosome. For example, the fusogenic agent, upon being exposed to the pH environment of an endosome, will cause a solubility or osmotic change substantial enough to increase the porosity of (preferably, to rupture) the endosomal membrane.

The fusogenic agent can be a polymer, preferably a polyamino chain, *e.g.*, polyethyleneimine (PEI). The PEI can be linear, branched, synthetic or natural. The PEI can be, *e.g.*, alkyl substituted PEI, or lipid substituted PEI.

In other embodiments, the fusogenic agent can be polyhistidine, polyimidazole, polypyridine, polypropyleneimine, mellitin, or a polyacetal substance, *e.g.*, a cationic polyacetal. In some embodiment, the fusogenic agent can have an alpha helical structure. The fusogenic agent can be a membrane disruptive agent, *e.g.*, mellittin.

Attorney's Docket No.: 14174-072W01

A fusogenic agent can have one or more of the following characteristics: (a) is at least 1Da in size; (b) has at least 10 charged groups, preferably between 10 and 5000 charged groups, more preferably between 50 and 1000 charged groups; (c) is present in the complex at a ratio of at least 1:1 fusogenic agent to carrier agent; (d) is present in the complex at a ratio of at least 1:1 fusogenic agent to condensing agent; (e) is present in the complex at a ratio of at least 1:1 fusogenic agent to targeting agent.

Other suitable fusogenic agents can be tested and identified by a skilled artisan. The ability of a compound to respond to, *e.g.*, change charge depending on, the pH environment can be tested by routine methods, *e.g.*, in a cellular assay. For example, a test compound is combined or contacted with a cell, and the cell is allowed to take up the test compound, *e.g.*, by endocytosis. An endosome preparation can then be made from the contacted cells and the endosome preparation compared to an endosome preparation from control cells. A change, *e.g.*, a decrease, in the endosome fraction from the contacted cell vs. the control cell indicates that the test compound can function as a fusogenic agent. Alternatively, the contacted cell and control cell can be evaluated, *e.g.*, by microscopy, *e.g.*, by light or electron microscopy, to determine a difference in endosome population in the cells. The test compound can be labeled. In another type of assay, a modular complex described herein is constructed using one or more test or putative fusogenic agents. The modular complex can be constructed using a labeled nucleic acid instead of the iRNA. The ability of the fusogenic agent to respond to, *e.g.*, change charge depending on, the pH environment, once the modular complex is taken up by the cell, can be evaluated, *e.g.*, by preparation of an endosome preparation, or by microscopy techniques, as described above. A two-step assay can also be performed, wherein a first assay evaluates the ability of a test compound alone to respond to, *e.g.*, change charge depending on, the pH environment; and a second assay evaluates the ability of a modular complex that includes the test compound to respond to, *e.g.*, change charge depending on, the pH environment.

Condensing agent

The condensing agent of a modular complex described herein can interact with (*e.g.*, attracts, holds, or binds to) an iRNA agent and act to (a) condense, *e.g.*, reduce the size or charge of the iRNA agent and/or (b) protect the iRNA agent, *e.g.*, protect the iRNA agent against

Attorney's Docket No.: 14174-072W01

degradation. The condensing agent can include a moiety, *e.g.*, a charged moiety, that can interact with a nucleic acid, *e.g.*, an iRNA agent, *e.g.*, by ionic interactions. The condensing agent would preferably be a charged polymer, *e.g.*, a polycationic chain. The condensing agent can be a polylysine (PLL), spermine, spermidine, polyamine, pseudopeptide-polyamine, peptidomimetic polyamine, dendrimer polyamine, arginine, amidine, protamine, cationic lipid, cationic porphyrin, quaternary salt of a polyamine, or an alpha helical peptide.

A condensing agent can have the following characteristics: (a) at least 1Da in size; (b) has at least 2 charged groups, preferably between 2 and 100 charged groups; (c) is present in the complex at a ratio of at least 1:1 condensing agent to carrier agent; (d) is present in the complex at a ratio of at least 1:1 condensing agent to fusogenic agent; (e) is present in the complex at a ratio of at least 1:1 condensing agent to targeting agent.

Other suitable condensing agents can be tested and identified by a skilled artisan, *e.g.*, by evaluating the ability of a test agent to interact with a nucleic acid, *e.g.*, an iRNA agent. The ability of a test agent to interact with a nucleic acid, *e.g.*, an iRNA agent, *e.g.*, to condense or protect the iRNA agent, can be evaluated by routine techniques. In one assay, a test agent is contacted with a nucleic acid, and the size and/or charge of the contacted nucleic acid is evaluated by a technique suitable to detect changes in molecular mass and/or charge. Such techniques include non-denaturing gel electrophoresis, immunological methods, *e.g.*, immunoprecipitation, gel filtration, ionic interaction chromatography, and the like. A test agent is identified as a condensing agent if it changes the mass and/or charge (preferably both) of the contacted nucleic acid, compared to a control. A two-step assay can also be performed, wherein a first assay evaluates the ability of a test compound alone to interact with, *e.g.*, bind to, *e.g.*, condense the charge and/or mass of, a nucleic acid; and a second assay evaluates the ability of a modular complex that includes the test compound to interact with, *e.g.*, bind to, *e.g.*, condense the charge and/or mass of, a nucleic acid.

Amphipathic Delivery Agents

An RNA, *e.g.*, an iRNA agent, described herein can be used with an amphipathic delivery conjugate or module, such as those described herein and those described in copending, co-owned United States Provisional Application Serial No. 60/455,050, filed on March 13, 2003, and

Attorney's Docket No.: 14174-072W01

International Application Serial No. PCT/US04/07070, filed March 8, 2004, which is hereby incorporated by reference.

In addition, the invention includes an iRNA agent described herein, e.g., a palindromic iRNA agent, an iRNA agent having a noncanonical pairing, an iRNA agent which targets a gene described herein, e.g., a gene active in the liver, an iRNA agent having a chemical modification described herein, e.g., a modification which enhances resistance to degradation, an iRNA agent having an architecture or structure described herein, an iRNA agent administered as described herein, or an iRNA agent formulated as described herein, combined with, associated with, and delivered by such an amphipathic delivery conjugate.

An amphipathic molecule is a molecule having a hydrophobic and a hydrophilic region. Such molecules can interact with (e.g., penetrate or disrupt) lipids, e.g., a lipid bilayer of a cell. As such, they can serve as delivery agent for an associated (e.g., bound) iRNA (e.g., an iRNA or sRNA described herein). A preferred amphipathic molecule to be used in the compositions described herein (e.g., the amphipathic iRNA constructs described herein) is a polymer. The polymer may have a secondary structure, e.g., a repeating secondary structure.

One example of an amphipathic polymer is an amphipathic polypeptide, e.g., a polypeptide having a secondary structure such that the polypeptide has a hydrophilic and a hydrophobic face. The design of amphipathic peptide structures (e.g., alpha-helical polypeptides) is routine to one of skill in the art. For example, the following references provide guidance:

Grell et al. (2001) "Protein design and folding: template trapping of self-assembled helical bundles" *J Pept Sci* 7(3):146-51; Chen et al. (2002) "Determination of stereochemistry stability coefficients of amino acid side-chains in an amphipathic alpha-helix" *J Pept Res* 59(1):18-33; Iwata et al. (1994) "Design and synthesis of amphipathic 3(10)-helical peptides and their interactions with phospholipid bilayers and ion channel formation" *J Biol Chem* 269(7):4928-33; Cornut et al. (1994) "The amphipathic alpha-helix concept. Application to the de novo design of ideally amphipathic Leu, Lys peptides with hemolytic activity higher than that of melittin" *FEBS Lett* 349(1):29-33; Negrete et al. (1998) "Deciphering the structural code for proteins: helical propensities in domain classes and statistical multiresidue information in alpha-helices," *Protein Sci* 7(6):1368-79.

Attorney's Docket No.: 14174-072W01

Another example of an amphipathic polymer is a polymer made up of two or more amphipathic subunits, e.g., two or more subunits containing cyclic moieties (e.g., a cyclic moiety having one or more hydrophilic groups and one or more hydrophobic groups). For example, the subunit may contain a steroid, e.g., cholic acid; or a aromatic moiety. Such moieties preferably
5 can exhibit atropisomerism, such that they can form opposing hydrophobic and hydrophilic faces when in a polymer structure.

The ability of a putative amphipathic molecule to interact with a lipid membrane, e.g., a cell membrane, can be tested by routine methods, e.g., in a cell free or cellular assay. For example, a test compound is combined or contacted with a synthetic lipid bilayer, a cellular
10 membrane fraction, or a cell, and the test compound is evaluated for its ability to interact with, penetrate or disrupt the lipid bilayer, cell membrane or cell. The test compound can be labeled in order to detect the interaction with the lipid bilayer, cell membrane or cell. In another type of assay, the test compound is linked to a reporter molecule or an iRNA agent (e.g., an iRNA or sRNA described herein) and the ability of the reporter molecule or iRNA agent to penetrate the
15 lipid bilayer, cell membrane or cell is evaluated. A two-step assay can also be performed, wherein a first assay evaluates the ability of a test compound alone to interact with a lipid bilayer, cell membrane or cell; and a second assay evaluates the ability of a construct (e.g., a construct described herein) that includes the test compound and a reporter or iRNA agent to interact with a lipid bilayer, cell membrane or cell.

20 An amphipathic polymer useful in the compositions described herein has at least 2, preferably at least 5, more preferably at least 10, 25, 50, 100, 200, 500, 1000, 2000, 50000 or more subunits (e.g., amino acids or cyclic subunits). A single amphipathic polymer can be linked to one or more, e.g., 2, 3, 5, 10 or more iRNA agents (e.g., iRNA or sRNA agents described herein). In some embodiments, an amphipathic polymer can contain both amino acid
25 and cyclic subunits, e.g., aromatic subunits.

The invention features a composition that includes an iRNA agent (e.g., an iRNA or sRNA described herein) in association with an amphipathic molecule. Such compositions may be referred to herein as "amphipathic iRNA constructs." Such compositions and constructs are useful in the delivery or targeting of iRNA agents, e.g., delivery or targeting of iRNA agents to a

Attorney's Docket No.: 14174-072W01

cell. While not wanting to be bound by theory, such compositions and constructs can increase the porosity of, e.g., can penetrate or disrupt, a lipid (e.g., a lipid bilayer of a cell), e.g., to allow entry of the iRNA agent into a cell.

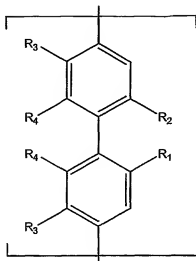
5 In one aspect, the invention relates to a composition comprising an iRNA agent (e.g., an iRNA or sRNA agent described herein) linked to an amphipathic molecule. The iRNA agent and the amphipathic molecule may be held in continuous contact with one another by either covalent or noncovalent linkages.

The amphipathic molecule of the composition or construct is preferably other than a phospholipid, e.g., other than a micelle, membrane or membrane fragment.

10 The amphipathic molecule of the composition or construct is preferably a polymer. The polymer may include two or more amphipathic subunits. One or more hydrophilic groups and one or more hydrophobic groups may be present on the polymer. The polymer may have a repeating secondary structure as well as a first face and a second face. The distribution of the hydrophilic groups and the hydrophobic groups along the repeating secondary structure can be
15 such that one face of the polymer is a hydrophilic face and the other face of the polymer is a hydrophobic face.

The amphipathic molecule can be a polypeptide, e.g., a polypeptide comprising an α -helical conformation as its secondary structure.

In one embodiment, the amphipathic polymer includes one or more subunits containing
20 one or more cyclic moiety (e.g., a cyclic moiety having one or more hydrophilic groups and/or one or more hydrophobic groups). In one embodiment, the polymer is a polymer of cyclic moieties such that the moieties have alternating hydrophobic and hydrophilic groups. For example, the subunit may contain a steroid, e.g., cholic acid. In another example, the subunit may contain an aromatic moiety. The aromatic moiety may be one that can exhibit
25 atropisomerism, e.g., a 2,2'-bis(substituted)-1-1'-binaphthyl or a 2,2'-bis(substituted) biphenyl. A subunit may include an aromatic moiety of Formula (M):



(M)

The invention features a composition that includes an iRNA agent (e.g., an iRNA or sRNA described herein) in association with an amphipathic molecule. Such compositions may be referred to herein as "amphipathic iRNA constructs." Such compositions and constructs are useful in the delivery or targeting of iRNA agents, e.g., delivery or targeting of iRNA agents to a cell. While not wanting to be bound by theory, such compositions and constructs can increase the porosity of, e.g., can penetrate or disrupt, a lipid (e.g., a lipid bilayer of a cell), e.g., to allow entry of the iRNA agent into a cell.

In one aspect, the invention relates to a composition comprising an iRNA agent (e.g., an iRNA or sRNA agent described herein) linked to an amphipathic molecule. The iRNA agent and the amphipathic molecule may be held in continuous contact with one another by either covalent or noncovalent linkages.

The amphipathic molecule of the composition or construct is preferably other than a phospholipid, e.g., other than a micelle, membrane or membrane fragment.

The amphipathic molecule of the composition or construct is preferably a polymer. The polymer may include two or more amphipathic subunits. One or more hydrophilic groups and

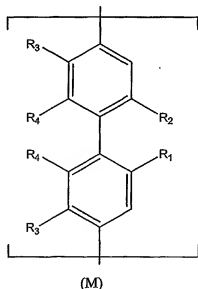
Attorney's Docket No.: 14174-072W01

one or more hydrophobic groups may be present on the polymer. The polymer may have a repeating secondary structure as well as a first face and a second face. The distribution of the hydrophilic groups and the hydrophobic groups along the repeating secondary structure can be such that one face of the polymer is a hydrophilic face and the other face of the polymer is a hydrophobic face.

The amphipathic molecule can be a polypeptide, e.g., a polypeptide comprising an α -helical conformation as its secondary structure.

In one embodiment, the amphipathic polymer includes one or more subunits containing one or more cyclic moiety (e.g., a cyclic moiety having one or more hydrophilic groups and/or one or more hydrophobic groups). In one embodiment, the polymer is a polymer of cyclic moieties such that the moieties have alternating hydrophobic and hydrophilic groups. For example, the subunit may contain a steroid, e.g., cholic acid. In another example, the subunit may contain an aromatic moiety. The aromatic moiety may be one that can exhibit atropisomerism, e.g., a 2,2'-bis(substituted)-1-1'-binaphthyl or a 2,2'-bis(substituted) biphenyl.

A subunit may include an aromatic moiety of Formula (M):



Attorney's Docket No.: 14174-072W01

Referring to Formula M, R_1 is C_1-C_{100} alkyl optionally substituted with aryl, alkenyl, alkynyl, alkoxy or halo and/or optionally inserted with O, S, alkenyl or alkynyl; C_1-C_{100} perfluoroalkyl; or OR_5 .

5 R_2 is hydroxy; nitro; sulfate; phosphate; phosphate ester; sulfonic acid; OR_6 ; or C_1-C_{100} alkyl optionally substituted with hydroxy, halo, nitro, aryl or alkyl sulfinyl, aryl or alkyl sulfonyl, sulfate, sulfonic acid, phosphate, phosphate ester, substituted or unsubstituted aryl, carboxyl, carboxylate, amino carbonyl, or alkoxycarbonyl, and/or optionally inserted with O, NH, S, S(O), SO_2 , alkenyl, or alkynyl.

10 R_3 is hydrogen, or when taken together with R_4 forms a fused phenyl ring.

R_4 is hydrogen, or when taken together with R_3 forms a fused phenyl ring.

R_5 is C_1-C_{100} alkyl optionally substituted with aryl, alkenyl, alkynyl, alkoxy or halo and/or optionally inserted with O, S, alkenyl or alkynyl; or C_1-C_{100} perfluoroalkyl; and R_6 is C_1-C_{100} alkyl optionally substituted with hydroxy, halo, nitro, aryl or alkyl sulfinyl, aryl or alkyl
15 sulfonyl, sulfate, sulfonic acid, phosphate, phosphate ester, substituted or unsubstituted aryl, carboxyl, carboxylate, amino carbonyl, or alkoxycarbonyl, and/or optionally inserted with O, NH, S, S(O), SO_2 , alkenyl, or alkynyl.

Increasing cellular uptake of dsRNAs

A method of the invention that can include the administration of an iRNA agent and a
20 drug that affects the uptake of the iRNA agent into the cell. The drug can be administered before, after, or at the same time that the iRNA agent is administered. The drug can be covalently linked to the iRNA agent. The drug can be, for example, a lipopolysaccharide, an activator of p38 MAP kinase, or an activator of NF- κ B. The drug can have a transient effect on the cell.

25 The drug can increase the uptake of the iRNA agent into the cell, for example, by disrupting the cell's cytoskeleton, *e.g.*, by disrupting the cell's microtubules, microfilaments, and/or intermediate filaments. The drug can be, for example, taxon, vincristine, vinblastine,

Attorney's Docket No.: 14174-072W01

cytochalasin, nocodazole, japlakinolide, latrunculin A, phalloidin, swinholide A, indanocine, or myoservin.

The drug can also increase the uptake of the iRNA agent into the cell by activating an inflammatory response, for example. Exemplary drug's that would have such an effect include
5 tumor necrosis factor alpha (TNFalpha), interleukin-1 beta, or gamma interferon.

iRNA conjugates

An iRNA agent can be coupled, *e.g.*, covalently coupled, to a second agent. For example, an iRNA agent used to treat a particular disorder can be coupled to a second therapeutic agent, *e.g.*, an agent other than the iRNA agent. The second therapeutic agent can be one which is
10 directed to the treatment of the same disorder. For example, in the case of an iRNA used to treat a disorder characterized by unwanted cell proliferation, *e.g.*, cancer, the iRNA agent can be coupled to a second agent which has an anti-cancer effect. For example, it can be coupled to an agent which stimulates the immune system, *e.g.*, a CpG motif, or more generally an agent that activates a toll-like receptor and/or increases the production of gamma interferon.

iRNA Production

An iRNA can be produced, *e.g.*, in bulk, by a variety of methods. Exemplary methods include: organic synthesis and RNA cleavage, *e.g.*, *in vitro* cleavage.

Organic Synthesis

An iRNA can be made by separately synthesizing each respective strand of a double-
20 stranded RNA molecule. The component strands can then be annealed.

A large bioreactor, *e.g.*, the OligoPilot II from Pharmacia Biotec AB (Uppsala Sweden), can be used to produce a large amount of a particular RNA strand for a given iRNA. The OligoPilotII reactor can efficiently couple a nucleotide using only a 1.5 molar excess of a phosphoramidite nucleotide. To make an RNA strand, ribonucleotides amidites are used.
25 Standard cycles of monomer addition can be used to synthesize the 21 to 23 nucleotide strand for the iRNA. Typically, the two complementary strands are produced separately and then annealed, *e.g.*, after release from the solid support and deprotection.

Attorney's Docket No.: 14174-072W01

Organic synthesis can be used to produce a discrete iRNA species. The complementary of the species to a particular target gene can be precisely specified. For example, the species may be complementary to a region that includes a polymorphism, *e.g.*, a single nucleotide polymorphism. Further the location of the polymorphism can be precisely defined. In some
5 embodiments, the polymorphism is located in an internal region, *e.g.*, at least 4, 5, 7, or 9 nucleotides from one or both of the termini.

dsRNA Cleavage

iRNAs can also be made by cleaving a larger ds iRNA. The cleavage can be mediated *in vitro* or *in vivo*. For example, to produce iRNAs by cleavage *in vitro*, the following method can
10 be used:

In vitro transcription. dsRNA is produced by transcribing a nucleic acid (DNA) segment in both directions. For example, the HiScribe™ RNAi transcription kit (New England Biolabs) provides a vector and a method for producing a dsRNA for a nucleic acid segment that is cloned into the vector at a position flanked on either side by a T7 promoter. Separate templates are
15 generated for T7 transcription of the two complementary strands for the dsRNA. The templates are transcribed *in vitro* by addition of T7 RNA polymerase and dsRNA is produced. Similar methods using PCR and/or other RNA polymerases (*e.g.*, T3 or SP6 polymerase) can also be used. In one embodiment, RNA generated by this method is carefully purified to remove endotoxins that may contaminate preparations of the recombinant enzymes.

20 *In vitro* cleavage. dsRNA is cleaved *in vitro* into iRNAs, for example, using a Dicer or comparable RNase III-based activity. For example, the dsRNA can be incubated in an *in vitro* extract from *Drosophila* or using purified components, *e.g.* a purified RNase or RISC complex (RNA-induced silencing complex). See, *e.g.*, Ketting *et al. Genes Dev* 2001 Oct 15;15(20):2654-9. and Hammond *Science* 2001 Aug 10;293(5532):1146-50.

25 dsRNA cleavage generally produces a plurality of iRNA species, each being a particular 21 to 23 nt fragment of a source dsRNA molecule. For example, iRNAs that include sequences complementary to overlapping regions and adjacent regions of a source dsRNA molecule may be present.

Attorney's Docket No.: 14174-072W01

Regardless of the method of synthesis, the iRNA preparation can be prepared in a solution (*e.g.*, an aqueous and/or organic solution) that is appropriate for formulation. For example, the iRNA preparation can be precipitated and redissolved in pure double-distilled water, and lyophilized. The dried iRNA can then be resuspended in a solution appropriate for the intended formulation process.

Synthesis of modified and nucleotide surrogate iRNA agents is discussed below.

FORMULATION

The iRNA agents described herein can be formulated for administration to a subject

For ease of exposition the formulations, compositions and methods in this section are discussed largely with regard to unmodified iRNA agents. It should be understood, however, that these formulations, compositions and methods can be practiced with other iRNA agents, *e.g.*, modified iRNA agents, and such practice is within the invention.

A formulated iRNA composition can assume a variety of states. In some examples, the composition is at least partially crystalline, uniformly crystalline, and/or anhydrous (*e.g.*, less than 80, 50, 30, 20, or 10% water). In another example, the iRNA is in an aqueous phase, *e.g.*, in a solution that includes water.

The aqueous phase or the crystalline compositions can, *e.g.*, be incorporated into a delivery vehicle, *e.g.*, a liposome (particularly for the aqueous phase) or a particle (*e.g.*, a microparticle as can be appropriate for a crystalline composition). Generally, the iRNA composition is formulated in a manner that is compatible with the intended method of administration (see, below).

In particular embodiments, the composition is prepared by at least one of the following methods: spray drying, lyophilization, vacuum drying, evaporation, fluid bed drying, or a combination of these techniques; or sonication with a lipid, freeze-drying, condensation and other self-assembly.

A iRNA preparation can be formulated in combination with another agent, *e.g.*, another therapeutic agent or an agent that stabilizes a iRNA, *e.g.*, a protein that complexes with iRNA to

Attorney's Docket No.: 14174-072W01

form an iRNP. Still other agents include chelators, *e.g.*, EDTA (*e.g.*, to remove divalent cations such as Mg^{2+}), salts, RNase inhibitors (*e.g.*, a broad specificity RNase inhibitor such as RNAsin) and so forth.

In one embodiment, the iRNA preparation includes another iRNA agent, *e.g.*, a second
5 iRNA that can mediated RNAi with respect to a second gene, or with respect to the same gene. Still other preparation can include at least 3, 5, ten, twenty, fifty, or a hundred or more different iRNA species. Such iRNAs can mediated RNAi with respect to a similar number of different genes.

In one embodiment, the iRNA preparation includes at least a second therapeutic agent
10 (*e.g.*, an agent other than an RNA or a DNA). For example, a iRNA composition for the treatment of a viral disease, *e.g.* HIV, might include a known antiviral agent (*e.g.*, a protease inhibitor or reverse transcriptase inhibitor). In another example, a iRNA composition for the treatment of a cancer might further comprise a chemotherapeutic agent.

Exemplary formulations are discussed below:

15 Liposomes

For ease of exposition the formulations, compositions and methods in this section are discussed largely with regard to unmodified iRNA agents. It should be understood, however, that these formulations, compositions and methods can be practiced with other iRNA agents, *e.g.*, modified iRNA s agents, and such practice is within the invention. An iRNA agent, *e.g.*, a
20 double-stranded iRNA agent, or sRNA agent, (*e.g.*, a precursor, *e.g.*, a larger iRNA agent which can be processed into a sRNA agent, or a DNA which encodes an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, or precursor thereof) preparation can be formulated for delivery in a membranous molecular assembly, *e.g.*, a liposome or a micelle. As used herein, the term "liposome" refers to a vesicle composed of amphiphilic lipids arranged in at least one
25 bilayer, *e.g.*, one bilayer or a plurality of bilayers. Liposomes include unilamellar and multilamellar vesicles that have a membrane formed from a lipophilic material and an aqueous interior. The aqueous portion contains the iRNA composition. The lipophilic material isolates the aqueous interior from an aqueous exterior, which typically does not include the iRNA

Attorney's Docket No.: 14174-072W01

composition, although in some examples, it may. Liposomes are useful for the transfer and delivery of active ingredients to the site of action. Because the liposomal membrane is structurally similar to biological membranes, when liposomes are applied to a tissue, the liposomal bilayer fuses with bilayer of the cellular membranes. As the merging of the liposome and cell progresses, the internal aqueous contents that include the iRNA are delivered into the cell where the iRNA can specifically bind to a target RNA and can mediate RNAi. In some cases the liposomes are also specifically targeted, *e.g.*, to direct the iRNA to particular cell types, *e.g.*, to cells of the liver, such as those described herein.

A liposome containing a iRNA can be prepared by a variety of methods.

In one example, the lipid component of a liposome is dissolved in a detergent so that micelles are formed with the lipid component. For example, the lipid component can be an amphipathic cationic lipid or lipid conjugate. The detergent can have a high critical micelle concentration and may be nonionic. Exemplary detergents include cholate, CHAPS, octylglucoside, deoxycholate, and lauroyl sarcosine. The iRNA preparation is then added to the micelles that include the lipid component. The cationic groups on the lipid interact with the iRNA and condense around the iRNA to form a liposome. After condensation, the detergent is removed, *e.g.*, by dialysis, to yield a liposomal preparation of iRNA.

If necessary a carrier compound that assists in condensation can be added during the condensation reaction, *e.g.*, by controlled addition. For example, the carrier compound can be a polymer other than a nucleic acid (*e.g.*, spermine or spermidine). pH can also adjusted to favor condensation.

Further description of methods for producing stable polynucleotide delivery vehicles, which incorporate a polynucleotide/cationic lipid complex as structural components of the delivery vehicle, are described in, *e.g.*, WO 96/37194. Liposome formation can also include one or more aspects of exemplary methods described in Felgner, P. L. *et al.*, *Proc. Natl. Acad. Sci.*, USA 8:7413-7417, 1987; U.S. Pat. No. 4,897,355; U.S. Pat. No. 5,171,678; Bangham, *et al. M. Mol. Biol.* 23:238, 1965; Olson, *et al. Biochim. Biophys. Acta* 557:9, 1979; Szoka, *et al. Proc. Natl. Acad. Sci.* 75: 4194, 1978; Mayhew, *et al. Biochim. Biophys. Acta* 775:169, 1984; Kim, *et al. Biochim. Biophys. Acta* 728:339, 1983; and Fukunaga, *et al. Endocrinol.* 115:757, 1984.

Attorney's Docket No.: 14174-072W01

Commonly used techniques for preparing lipid aggregates of appropriate size for use as delivery vehicles include sonication and freeze-thaw plus extrusion (see, e.g., Mayer, *et al. Biochim. Biophys. Acta* 858:161, 1986). Microfluidization can be used when consistently small (50 to 200 nm) and relatively uniform aggregates are desired (Mayhew, *et al. Biochim. Biophys. Acta* 775:169, 1984). These methods are readily adapted to packaging iRNA preparations into liposomes.

Liposomes that are pH-sensitive or negatively-charged, entrap nucleic acid molecules rather than complex with them. Since both the nucleic acid molecules and the lipid are similarly charged, repulsion rather than complex formation occurs. Nevertheless, some nucleic acid molecules are entrapped within the aqueous interior of these liposomes. pH-sensitive liposomes have been used to deliver DNA encoding the thymidine kinase gene to cell monolayers in culture. Expression of the exogenous gene was detected in the target cells (Zhou *et al., Journal of Controlled Release*, 19, (1992) 269-274).

One major type of liposomal composition includes phospholipids other than naturally-derived phosphatidylcholine. Neutral liposome compositions, for example, can be formed from dimyristoyl phosphatidylcholine (DMPC) or dipalmitoyl phosphatidylcholine (DPPC). Anionic liposome compositions generally are formed from dimyristoyl phosphatidylglycerol, while anionic fusogenic liposomes are formed primarily from dioleoyl phosphatidylethanolamine (DOPE). Another type of liposomal composition is formed from phosphatidylcholine (PC) such as, for example, soybean PC, and egg PC. Another type is formed from mixtures of phospholipid and/or phosphatidylcholine and/or cholesterol.

Examples of other methods to introduce liposomes into cells *in vitro* and *in vivo* include U.S. Pat. No. 5,283,185; U.S. Pat. No. 5,171,678; WO 94/00569; WO 93/24640; WO 91/16024; Felgner, *J. Biol. Chem.* 269:2550, 1994; Nabel, *Proc. Natl. Acad. Sci.* 90:11307, 1993; Nabel, *Human Gene Ther.* 3:649, 1992; Gershon, *Biochem.* 32:7143, 1993; and Strauss *EMBO J.* 11:417, 1992.

In one embodiment, cationic liposomes are used. Cationic liposomes possess the advantage of being able to fuse to the cell membrane. Non-cationic liposomes, although not able

Attorney's Docket No.: 14174-072W01

to fuse as efficiently with the plasma membrane, are taken up by macrophages *in vivo* and can be used to deliver iRNAs to macrophages.

Further advantages of liposomes include: liposomes obtained from natural phospholipids are biocompatible and biodegradable; liposomes can incorporate a wide range of water and lipid soluble drugs; liposomes can protect encapsulated iRNAs in their internal compartments from metabolism and degradation (Rosoff, in "Pharmaceutical Dosage Forms," Lieberman, Rieger and Banker (Eds.), 1988, volume 1, p. 245). Important considerations in the preparation of liposome formulations are the lipid surface charge, vesicle size and the aqueous volume of the liposomes.

A positively charged synthetic cationic lipid, N-[1-(2,3-dioleoyloxy)propyl]-N,N,N-trimethylammonium chloride (DOTMA) can be used to form small liposomes that interact spontaneously with nucleic acid to form lipid-nucleic acid complexes which are capable of fusing with the negatively charged lipids of the cell membranes of tissue culture cells, resulting in delivery of iRNA (see, *e.g.*, Felgner, P. L. *et al.*, Proc. Natl. Acad. Sci., USA 8:7413-7417, 1987 and U.S. Pat. No. 4,897,355 for a description of DOTMA and its use with DNA).

A DOTMA analogue, 1,2-bis(oleoyloxy)-3-(trimethylammonia)propane (DOTAP) can be used in combination with a phospholipid to form DNA-complexing vesicles. Lipofectin™ (Bethesda Research Laboratories, Gaithersburg, Md.) is an effective agent for the delivery of highly anionic nucleic acids into living tissue culture cells that comprise positively charged DOTMA liposomes which interact spontaneously with negatively charged polynucleotides to form complexes. When enough positively charged liposomes are used, the net charge on the resulting complexes is also positive. Positively charged complexes prepared in this way spontaneously attach to negatively charged cell surfaces, fuse with the plasma membrane, and efficiently deliver functional nucleic acids into, for example, tissue culture cells. Another commercially available cationic lipid, 1,2-bis(oleoyloxy)-3,3-(trimethylammonia)propane ("DOTAP") (Boehringer Mannheim, Indianapolis, Indiana) differs from DOTMA in that the oleoyl moieties are linked by ester, rather than ether linkages.

Other reported cationic lipid compounds include those that have been conjugated to a variety of moieties including, for example, carboxyspermine which has been conjugated to one of two types of lipids and includes compounds such as 5-carboxyspermylglycine

Attorney's Docket No.: 14174-072W01

dioctaoeoylamide ("DOGS") (Transfectam™, Promega, Madison, Wisconsin) and dipalmitoylphosphatidylethanolamine 5-carboxyspermoyl-amide ("DPPES") (see, e.g., U.S. Pat. No. 5,171,678).

Another cationic lipid conjugate includes derivatization of the lipid with cholesterol ("DC-Chol") which has been formulated into liposomes in combination with DOPE (See, Gao, X. and Huang, L., *Biochim. Biophys. Res. Commun.* 179:280, 1991). Lipopolylysine, made by conjugating polylysine to DOPE, has been reported to be effective for transfection in the presence of serum (Zhou, X. *et al.*, *Biochim. Biophys. Acta* 1065:8, 1991). For certain cell lines, these liposomes containing conjugated cationic lipids, are said to exhibit lower toxicity and provide more efficient transfection than the DOTMA-containing compositions. Other commercially available cationic lipid products include DMRIE and DMRIE-HP (Vical, La Jolla, California) and Lipofectamine (DOSPA) (Life Technology, Inc., Gaithersburg, Maryland). Other cationic lipids suitable for the delivery of oligonucleotides are described in WO 98/39359 and WO 96/37194.

Liposomal formulations are particularly suited for topical administration, liposomes present several advantages over other formulations. Such advantages include reduced side effects related to high systemic absorption of the administered drug, increased accumulation of the administered drug at the desired target, and the ability to administer iRNA, into the skin. In some implementations, liposomes are used for delivering iRNA to epidermal cells and also to enhance the penetration of iRNA into dermal tissues, e.g., into skin. For example, the liposomes can be applied topically. Topical delivery of drugs formulated as liposomes to the skin has been documented (see, e.g., Weiner *et al.*, *Journal of Drug Targeting*, 1992, vol. 2,405-410 and du Plessis *et al.*, *Antiviral Research*, 18, 1992, 259-265; Mannino, R. J. and Fould-Fogerite, S., *Biotechniques* 6:682-690, 1988; Itani, T. *et al.* *Gene* 56:267-276. 1987; Nicolau, C. *et al.* *Meth. Enz.* 149:157-176, 1987; Straubinger, R. M. and Papahadjopoulos, D. *Meth. Enz.* 101:512-527, 1983; Wang, C. Y. and Huang, L., *Proc. Natl. Acad. Sci. USA* 84:7851-7855, 1987).

Non-ionic liposomal systems have also been examined to determine their utility in the delivery of drugs to the skin, in particular systems comprising non-ionic surfactant and cholesterol. Non-ionic liposomal formulations comprising Novasome I (glyceryl

Attorney's Docket No.: 14174-072W01

dilaurate/cholesterol/polyoxyethylene-10-stearyl ether) and Novasome II (glyceryl distearate/cholesterol/polyoxyethylene-10-stearyl ether) were used to deliver a drug into the dermis of mouse skin. Such formulations with iRNA are useful for treating a dermatological disorder.

5 Liposomes that include iRNA can be made highly deformable. Such deformability can enable the liposomes to penetrate through pore that are smaller than the average radius of the liposome. For example, transfersomes are a type of deformable liposomes. Transfersomes can be made by adding surface edge activators, usually surfactants, to a standard liposomal composition. Transfersomes that include iRNA can be delivered, for example, subcutaneously by infection in order to deliver iRNA to keratinocytes in the skin. In order to cross intact
10 mammalian skin, lipid vesicles must pass through a series of fine pores, each with a diameter less than 50 nm, under the influence of a suitable transdermal gradient. In addition, due to the lipid properties, these transfersomes can be self-optimizing (adaptive to the shape of pores, e.g., in the skin), self-repairing, and can frequently reach their targets without fragmenting, and often self-loading. The iRNA agents can include an RRMS tethered to a moiety which improves
15 association with a liposome.

Surfactants

For ease of exposition the formulations, compositions and methods in this section are discussed largely with regard to unmodified iRNA agents. It should be understood, however, that these formulations, compositions and methods can be practiced with other iRNA agents,
20 e.g., modified iRNA agents, and such practice is within the invention. Surfactants find wide application in formulations such as emulsions (including microemulsions) and liposomes (see above). iRNA (or a precursor, e.g., a larger dsRNA which can be processed into a iRNA, or a DNA which encodes a iRNA or precursor) compositions can include a surfactant. In one embodiment, the iRNA is formulated as an emulsion that includes a surfactant. The most
25 common way of classifying and ranking the properties of the many different types of surfactants, both natural and synthetic, is by the use of the hydrophile/lipophile balance (HLB). The nature of the hydrophilic group provides the most useful means for categorizing the different surfactants used in formulations (Rieger, in "Pharmaceutical Dosage Forms," Marcel Dekker, Inc., New York, NY, 1988, p. 285).

Attorney's Docket No.: 14174-072W01

If the surfactant molecule is not ionized, it is classified as a nonionic surfactant.

Nonionic surfactants find wide application in pharmaceutical products and are usable over a wide range of pH values. In general their HLB values range from 2 to about 18 depending on their structure. Nonionic surfactants include nonionic esters such as ethylene glycol esters, propylene glycol esters, glyceryl esters, polyglyceryl esters, sorbitan esters, sucrose esters, and ethoxylated esters. Nonionic alkanolamides and ethers such as fatty alcohol ethoxylates, propoxylated alcohols, and ethoxylated/propoxylated block polymers are also included in this class. The polyoxyethylene surfactants are the most popular members of the nonionic surfactant class.

If the surfactant molecule carries a negative charge when it is dissolved or dispersed in water, the surfactant is classified as anionic. Anionic surfactants include carboxylates such as soaps, acyl lactylates, acyl amides of amino acids, esters of sulfuric acid such as alkyl sulfates and ethoxylated alkyl sulfates, sulfonates such as alkyl benzene sulfonates, acyl isethionates, acyl taurates and sulfosuccinates, and phosphates. The most important members of the anionic surfactant class are the alkyl sulfates and the soaps.

If the surfactant molecule carries a positive charge when it is dissolved or dispersed in water, the surfactant is classified as cationic. Cationic surfactants include quaternary ammonium salts and ethoxylated amines. The quaternary ammonium salts are the most used members of this class.

If the surfactant molecule has the ability to carry either a positive or negative charge, the surfactant is classified as amphoteric. Amphoteric surfactants include acrylic acid derivatives, substituted alkylamides, N-alkylbetaines and phosphatides.

The use of surfactants in drug products, formulations and in emulsions has been reviewed (Rieger, in "Pharmaceutical Dosage Forms," Marcel Dekker, Inc., New York, NY, 1988, p. 285).

Micelles and other Membranous Formulations

For ease of exposition the micelles and other formulations, compositions and methods in this section are discussed largely with regard to unmodified iRNA agents. It should be understood, however, that these micelles and other formulations, compositions and methods can

Attorney's Docket No.: 14174-072W01

be practiced with other iRNA agents, *e.g.*, modified iRNA agents, and such practice is within the invention. The iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, (*e.g.*, a precursor, *e.g.*, a larger iRNA agent which can be processed into a sRNA agent, or a DNA which encodes an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, or precursor thereof)) composition can be provided as a micellar formulation. "Micelles" are defined herein as a particular type of molecular assembly in which amphipathic molecules are arranged in a spherical structure such that all the hydrophobic portions of the molecules are directed inward, leaving the hydrophilic portions in contact with the surrounding aqueous phase. The converse arrangement exists if the environment is hydrophobic.

A mixed micellar formulation suitable for delivery through transdermal membranes may be prepared by mixing an aqueous solution of the iRNA composition, an alkali metal C₈ to C₂₂ alkyl sulphate, and a micelle forming compounds. Exemplary micelle forming compounds include lecithin, hyaluronic acid, pharmaceutically acceptable salts of hyaluronic acid, glycolic acid, lactic acid, chamomile extract, cucumber extract, oleic acid, linoleic acid, linolenic acid, monoolein, monooleates, monolaurates, borage oil, evening of primrose oil, menthol, trihydroxy oxo cholanyl glycine and pharmaceutically acceptable salts thereof, glycerin, polyglycerin, lysine, polylysine, triolein, polyoxyethylene ethers and analogues thereof, polidocanol alkyl ethers and analogues thereof, chenodeoxycholate, deoxycholate, and mixtures thereof. The micelle forming compounds may be added at the same time or after addition of the alkali metal alkyl sulphate. Mixed micelles will form with substantially any kind of mixing of the ingredients but vigorous mixing is preferred in order to provide smaller size micelles.

In one method a first micellar composition is prepared which contains the iRNA composition and at least the alkali metal alkyl sulphate. The first micellar composition is then mixed with at least three micelle forming compounds to form a mixed micellar composition. In another method, the micellar composition is prepared by mixing the iRNA composition, the alkali metal alkyl sulphate and at least one of the micelle forming compounds, followed by addition of the remaining micelle forming compounds, with vigorous mixing.

Phenol and/or m-cresol may be added to the mixed micellar composition to stabilize the formulation and protect against bacterial growth. Alternatively, phenol and/or m-cresol may be

Attorney's Docket No.: 14174-072W01

added with the micelle forming ingredients. An isotonic agent such as glycerin may also be added after formation of the mixed micellar composition.

For delivery of the micellar formulation as a spray, the formulation can be put into an aerosol dispenser and the dispenser is charged with a propellant. The propellant, which is under pressure, is in liquid form in the dispenser. The ratios of the ingredients are adjusted so that the aqueous and propellant phases become one, *i.e.* there is one phase. If there are two phases, it is necessary to shake the dispenser prior to dispensing a portion of the contents, *e.g.* through a metered valve. The dispensed dose of pharmaceutical agent is propelled from the metered valve in a fine spray.

The preferred propellants are hydrogen-containing chlorofluorocarbons, hydrogen-containing fluorocarbons, dimethyl ether and diethyl ether. Even more preferred is HFA 134a (1,1,1,2 tetrafluoroethane).

The specific concentrations of the essential ingredients can be determined by relatively straightforward experimentation. For absorption through the oral cavities, it is often desirable to increase, *e.g.* at least double or triple, the dosage for through injection or administration through the gastrointestinal tract.

The iRNA agents can include an RRMS tethered to a moiety which improves association with a micelle or other membranous formulation.

Particles

For ease of exposition the particles, formulations, compositions and methods in this section are discussed largely with regard to unmodified iRNA agents. It should be understood, however, that these particles, formulations, compositions and methods can be practiced with other iRNA agents, *e.g.*, modified iRNA agents, and such practice is within the invention. In another embodiment, an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, (*e.g.*, a precursor, *e.g.*, a larger iRNA agent which can be processed into a sRNA agent, or a DNA which encodes an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, or precursor thereof) preparations may be incorporated into a particle, *e.g.*, a microparticle. Microparticles can be produced by spray-drying, but may also be produced by other methods including

Attorney's Docket No.: 14174-072W01

lyophilization, evaporation, fluid bed drying, vacuum drying, or a combination of these techniques. See below for further description.

- Sustained-Release Formulations. An iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, (*e.g.*, a precursor, *e.g.*, a larger iRNA agent which can be processed into a sRNA agent, or a DNA which encodes an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, or precursor thereof) described herein can be formulated for controlled, *e.g.*, slow release. Controlled release can be achieved by disposing the iRNA within a structure or substance which impedes its release. *E.g.*, iRNA can be disposed within a porous matrix or in an erodable matrix, either of which allow release of the iRNA over a period of time.
- 10 Polymeric particles, *e.g.*, polymeric in microparticles can be used as a sustained-release reservoir of iRNA that is taken up by cells only released from the microparticle through biodegradation. The polymeric particles in this embodiment should therefore be large enough to preclude phagocytosis (*e.g.*, larger than 10 μm and preferably larger than 20 μm). Such particles can be produced by the same methods to make smaller particles, but with less vigorous mixing of
- 15 the first and second emulsions. That is to say, a lower homogenization speed, vortex mixing speed, or sonication setting can be used to obtain particles having a diameter around 100 μm rather than 10 μm . The time of mixing also can be altered.

- Larger microparticles can be formulated as a suspension, a powder, or an implantable solid, to be delivered by intramuscular, subcutaneous, intradermal, intravenous, or intraperitoneal
- 20 injection; via inhalation (intranasal or intrapulmonary); orally; or by implantation. These particles are useful for delivery of any iRNA when slow release over a relatively long term is desired. The rate of degradation, and consequently of release, varies with the polymeric formulation.

- Microparticles preferably include pores, voids, hollows, defects or other interstitial
- 25 spaces that allow the fluid suspension medium to freely permeate or perfuse the particulate boundary. For example, the perforated microstructures can be used to form hollow, porous spray dried microspheres.

Attorney's Docket No.: 14174-072W01

Polymeric particles containing iRNA (e.g., a sRNA) can be made using a double emulsion technique, for instance. First, the polymer is dissolved in an organic solvent. A preferred polymer is polylactic-co-glycolic acid (PLGA), with a lactic/glycolic acid weight ratio of 65:35, 50:50, or 75:25. Next, a sample of nucleic acid suspended in aqueous solution is added to the polymer solution and the two solutions are mixed to form a first emulsion. The solutions can be mixed by vortexing or shaking, and in a preferred method, the mixture can be sonicated. Most preferable is any method by which the nucleic acid receives the least amount of damage in the form of nicking, shearing, or degradation, while still allowing the formation of an appropriate emulsion. For example, acceptable results can be obtained with a Vibra-cell model VC-250 sonicator with a 1/8" microtip probe, at setting #3.

Spray-Drying. An iRNA agent, e.g., a double-stranded iRNA agent, or sRNA agent, (e.g., a precursor, e.g., a larger iRNA agent which can be processed into a sRNA agent, or a DNA which encodes an iRNA agent, e.g., a double-stranded iRNA agent, or sRNA agent, or precursor thereof) can be prepared by spray drying. Spray dried iRNA can be administered to a subject or be subjected to further formulation. A pharmaceutical composition of iRNA can be prepared by spray drying a homogeneous aqueous mixture that includes a iRNA under conditions sufficient to provide a dispersible powdered composition, e.g., a pharmaceutical composition. The material for spray drying can also include one or more of: a pharmaceutically acceptable excipient, or a dispersibility-enhancing amount of a physiologically acceptable, water-soluble protein. The spray-dried product can be a dispersible powder that includes the iRNA.

Spray drying is a process that converts a liquid or slurry material to a dried particulate form. Spray drying can be used to provide powdered material for various administrative routes including inhalation. See, for example, M. Sacchetti and M. M. Van Oort in: *Inhalation Aerosols: Physical and Biological Basis for Therapy*, A. J. Hickey, ed. Marcel Dekkar, New York, 1996.

Spray drying can include atomizing a solution, emulsion, or suspension to form a fine mist of droplets and drying the droplets. The mist can be projected into a drying chamber (e.g., a vessel, tank, tubing, or coil) where it contacts a drying gas. The mist can include solid or liquid pore forming agents. The solvent and pore forming agents evaporate from the droplets into the

Attorney's Docket No.: 14174-072W01

drying gas to solidify the droplets, simultaneously forming pores throughout the solid. The solid (typically in a powder, particulate form) then is separated from the drying gas and collected.

Spray drying includes bringing together a highly dispersed liquid, and a sufficient volume of air (e.g., hot air) to produce evaporation and drying of the liquid droplets. The preparation to
5 be spray dried can be any solution, course suspension, slurry, colloidal-dispersion, or paste that may be atomized using the selected spray drying apparatus. Typically, the feed is sprayed into a current of warm filtered air that evaporates the solvent and conveys the dried product to a collector. The spent air is then exhausted with the solvent. Several different types of apparatus may be used to provide the desired product. For example, commercial spray dryers manufactured
10 by Buchi Ltd. or Niro Corp. can effectively produce particles of desired size.

Spray-dried powdered particles can be approximately spherical in shape, nearly uniform in size and frequently hollow. There may be some degree of irregularity in shape depending upon the incorporated medicament and the spray drying conditions. In many instances the dispersion stability of spray-dried microspheres appears to be more effective if an inflating agent
15 (or blowing agent) is used in their production. Particularly preferred embodiments may comprise an emulsion with an inflating agent as the disperse or continuous phase (the other phase being aqueous in nature). An inflating agent is preferably dispersed with a surfactant solution, using, for instance, a commercially available microfluidizer at a pressure of about 5000 to 15,000 psi. This process forms an emulsion, preferably stabilized by an incorporated surfactant, typically
20 comprising submicron droplets of water immiscible blowing agent dispersed in an aqueous continuous phase. The formation of such dispersions using this and other techniques are common and well known to those in the art. The blowing agent is preferably a fluorinated compound (e.g. perfluorohexane, perfluorooctyl bromide, perfluorodecalin, perfluorobutyl ethane) which vaporizes during the spray-drying process, leaving behind generally hollow, porous
25 aerodynamically light microspheres. As will be discussed in more detail below, other suitable blowing agents include chloroform, freons, and hydrocarbons. Nitrogen gas and carbon dioxide are also contemplated as a suitable blowing agent.

Although the perforated microstructures are preferably formed using a blowing agent as described above, it will be appreciated that, in some instances, no blowing agent is required and

Attorney's Docket No.: 14174-072W01

an aqueous dispersion of the medicament and surfactant(s) are spray dried directly. In such cases, the formulation may be amenable to process conditions (e.g., elevated temperatures) that generally lead to the formation of hollow, relatively porous microparticles. Moreover, the medicament may possess special physicochemical properties (e.g., high crystallinity, elevated melting temperature, surface activity, etc.) that make it particularly suitable for use in such techniques.

The perforated microstructures may optionally be associated with, or comprise, one or more surfactants. Moreover, miscible surfactants may optionally be combined with the suspension medium liquid phase. It will be appreciated by those skilled in the art that the use of surfactants may further increase dispersion stability, simplify formulation procedures or increase bioavailability upon administration. Of course combinations of surfactants, including the use of one or more in the liquid phase and one or more associated with the perforated microstructures are contemplated as being within the scope of the invention. By "associated with or comprise" it is meant that the structural matrix or perforated microstructure may incorporate, adsorb, absorb, be coated with or be formed by the surfactant.

Surfactants suitable for use include any compound or composition that aids in the formation and maintenance of the stabilized respiratory dispersions by forming a layer at the interface between the structural matrix and the suspension medium. The surfactant may comprise a single compound or any combination of compounds, such as in the case of co-surfactants. Particularly preferred surfactants are substantially insoluble in the propellant, nonfluorinated, and selected from the group consisting of saturated and unsaturated lipids, nonionic detergents, nonionic block copolymers, ionic surfactants, and combinations of such agents. It should be emphasized that, in addition to the aforementioned surfactants, suitable (*i.e.* biocompatible) fluorinated surfactants are compatible with the teachings herein and may be used to provide the desired stabilized preparations.

Lipids, including phospholipids, from both natural and synthetic sources may be used in varying concentrations to form a structural matrix. Generally, compatible lipids comprise those that have a gel to liquid crystal phase transition greater than about 40° C. Preferably, the incorporated lipids are relatively long chain (*i.e.* C₆-C₂₂) saturated lipids and more preferably

Attorney's Docket No.: 14174-072W01

comprise phospholipids. Exemplary phospholipids useful in the disclosed stabilized preparations comprise egg phosphatidylcholine, dilauroylphosphatidylcholine, dioleoylphosphatidylcholine, dipalmitoylphosphatidyl-choline, distearylphosphatidylcholine, short-chain phosphatidylcholines, phosphatidylethanolamine, dioleoylphosphatidylethanolamine, 5 phosphatidylserine, phosphatidylglycerol, phosphatidylinositol, glycolipids, ganglioside GM1, sphingomyelin, phosphatidic acid, cardiolipin; lipids bearing polymer chains such as, polyethylene glycol, chitin, hyaluronic acid, or polyvinylpyrrolidone; lipids bearing sulfonated mono-, di-, and polysaccharides; fatty acids such as palmitic acid, stearic acid, and oleic acid; cholesterol, cholesterol esters, and cholesterol hemisuccinate. Due to their excellent 10 biocompatibility characteristics, phospholipids and combinations of phospholipids and poloxamers are particularly suitable for use in the stabilized dispersions disclosed herein.

Compatible nonionic detergents comprise: sorbitan esters including sorbitan trioleate (Spans™ 85), sorbitan sesquioleate, sorbitan monooleate, sorbitan monolaurate, polyoxyethylene (20) sorbitan monolaurate, and polyoxyethylene (20) sorbitan monooleate, oleyl polyoxyethylene 15 (2) ether, stearyl polyoxyethylene (2) ether, lauryl polyoxyethylene (4) ether, glycerol esters, and sucrose esters. Other suitable nonionic detergents can be easily identified using McCutcheon's Emulsifiers and Detergents (McPublishing Co., Glen Rock, N.J.). Preferred block copolymers include diblock and triblock copolymers of polyoxyethylene and polyoxypropylene, including poloxamer 188 (Pluronic.RTM. F68), poloxamer 407 (Pluronic.RTM. F-127), and poloxamer 20 338. Ionic surfactants such as sodium sulfosuccinate, and fatty acid soaps may also be utilized. In preferred embodiments, the microstructures may comprise oleic acid or its alkali salt.

In addition to the aforementioned surfactants, cationic surfactants or lipids are preferred especially in the case of delivery of an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, (*e.g.*, a precursor, *e.g.*, a larger iRNA agent which can be processed into a sRNA 25 agent, or a DNA which encodes an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, or precursor thereof). Examples of suitable cationic lipids include: DOTMA, N-[-(2,3-dioleoyloxy)propyl]-N,N,N-trimethylammonium-chloride; DOTAP, 1,2-dioleoyloxy-3-(trimethylammonio)propane; and DOTB, 1,2-dioleoyl-3-(4'-trimethylammonio)butanoyl-sn-glycerol. Polycationic amino acids such as polylysine, and polyarginine are also contemplated.

Attorney's Docket No.: 14174-072W01

For the spraying process, such spraying methods as rotary atomization, pressure atomization and two-fluid atomization can be used. Examples of the devices used in these processes include "Parubisu [phonetic rendering] Mini-Spray GA-32" and "Parubisu Spray Drier DL-41", manufactured by Yamato Chemical Co., or "Spray Drier CL-8," "Spray Drier L-8,"
5 "Spray Drier FL-12," "Spray Drier FL-16" or "Spray Drier FL-20," manufactured by Okawara Kakoki Co., can be used for the method of spraying using rotary-disk atomizer.

While no particular restrictions are placed on the gas used to dry the sprayed material, it is recommended to use air, nitrogen gas or an inert gas. The temperature of the inlet of the gas used to dry the sprayed materials such that it does not cause heat deactivation of the sprayed
10 material. The range of temperatures may vary between about 50°C to about 200°C, preferably between about 50°C and 100°C. The temperature of the outlet gas used to dry the sprayed material, may vary between about 0°C and about 150°C, preferably between 0°C and 90°C, and even more preferably between 0°C and 60°C.

The spray drying is done under conditions that result in substantially amorphous powder
15 of homogeneous constitution having a particle size that is respirable, a low moisture content and flow characteristics that allow for ready aerosolization. Preferably the particle size of the resulting powder is such that more than about 98% of the mass is in particles having a diameter of about 10 μm or less with about 90% of the mass being in particles having a diameter less than 5 μm . Alternatively, about 95% of the mass will have particles with a diameter of less than 10
20 μm with about 80% of the mass of the particles having a diameter of less than 5 μm .

The dispersible pharmaceutical-based dry powders that include the iRNA preparation may optionally be combined with pharmaceutical carriers or excipients which are suitable for respiratory and pulmonary administration. Such carriers may serve simply as bulking agents
when it is desired to reduce the iRNA concentration in the powder which is being delivered to a
25 patient, but may also serve to enhance the stability of the iRNA compositions and to improve the dispersibility of the powder within a powder dispersion device in order to provide more efficient and reproducible delivery of the iRNA and to improve handling characteristics of the iRNA such as flowability and consistency to facilitate manufacturing and powder filling.

Attorney's Docket No.: 14174-072W01

Such carrier materials may be combined with the drug prior to spray drying, *i.e.*, by adding the carrier material to the purified bulk solution. In that way, the carrier particles will be formed simultaneously with the drug particles to produce a homogeneous powder. Alternatively, the carriers may be separately prepared in a dry powder form and combined with the dry powder drug by blending. The powder carriers will usually be crystalline (to avoid water absorption), but might in some cases be amorphous or mixtures of crystalline and amorphous. The size of the carrier particles may be selected to improve the flowability of the drug powder, typically being in the range from 25 μm to 100 μm . A preferred carrier material is crystalline lactose having a size in the above-stated range.

Powders prepared by any of the above methods will be collected from the spray dryer in a conventional manner for subsequent use. For use as pharmaceuticals and other purposes, it will frequently be desirable to disrupt any agglomerates which may have formed by screening or other conventional techniques. For pharmaceutical uses, the dry powder formulations will usually be measured into a single dose, and the single dose sealed into a package. Such packages are particularly useful for dispersion in dry powder inhalers, as described in detail below. Alternatively, the powders may be packaged in multiple-dose containers.

Methods for spray drying hydrophobic and other drugs and components are described in U.S. Pat. Nos. 5,000,888; 5,026,550; 4,670,419, 4,540,602; and 4,486,435. Bloch and Speison (1983) Pharm. Acta Helv 58:14-22 teaches spray drying of hydrochlorothiazide and chlorthalidone (lipophilic drugs) and a hydrophilic adjuvant (pentaerythritol) in azeotropic solvents of dioxane-water and 2-ethoxyethanol-water. A number of Japanese Patent application Abstracts relate to spray drying of hydrophilic-hydrophobic product combinations, including JP 806766; JP 7242568; JP 7101884; JP 7101883; JP 71018982; JP 7101881; and JP 4036233. Other foreign patent publications relevant to spray drying hydrophilic-hydrophobic product combinations include FR 2594693; DE 2209477; and WO 88/07870.

LYOPHILIZATION.

An iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, (*e.g.*, a precursor, *e.g.*, a larger iRNA agent which can be processed into a sRNA agent, or a DNA which encodes an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, or precursor thereof)

Attorney's Docket No.: 14174-072W01

preparation can be made by lyophilization. Lyophilization is a freeze-drying process in which water is sublimed from the composition after it is frozen. The particular advantage associated with the lyophilization process is that biologicals and pharmaceuticals that are relatively unstable in an aqueous solution can be dried without elevated temperatures (thereby eliminating the adverse thermal effects), and then stored in a dry state where there are few stability problems. With respect to the instant invention such techniques are particularly compatible with the incorporation of nucleic acids in perforated microstructures without compromising physiological activity. Methods for providing lyophilized particulates are known to those of skill in the art and it would clearly not require undue experimentation to provide dispersion compatible microstructures in accordance with the teachings herein. Accordingly, to the extent that lyophilization processes may be used to provide microstructures having the desired porosity and size, they are conformance with the teachings herein and are expressly contemplated as being within the scope of the instant invention.

Targeting

For ease of exposition the formulations, compositions and methods in this section are discussed largely with regard to unmodified iRNAs. It should be understood, however, that these formulations, compositions and methods can be practiced with other iRNA agents, *e.g.*, modified iRNA agents, and such practice is within the invention.

In some embodiments, an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, (*e.g.*, a precursor, *e.g.*, a larger iRNA agent which can be processed into a sRNA agent, or a DNA which encodes an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, or precursor thereof) is targeted to a particular cell. For example, a liposome or particle or other structure that includes a iRNA can also include a targeting moiety that recognizes a specific molecule on a target cell. The targeting moiety can be a molecule with a specific affinity for a target cell. Targeting moieties can include antibodies directed against a protein found on the surface of a target cell, or the ligand or a receptor-binding portion of a ligand for a molecule found on the surface of a target cell. For example, the targeting moiety can recognize a cancer-specific antigen of the liver or a viral antigen, thus delivering the iRNA to a cancer cell or a virus-infected cell. Exemplary targeting moieties include antibodies (such as IgM, IgG, IgA,

Attorney's Docket No.: 14174-072W01

IgD, and the like, or a functional portions thereof), ligands for cell surface receptors (*e.g.*, ectodomains thereof).

An antigen, such as a-feto protein, can be used to target an iRNA to a liver cell.

In one embodiment, the targeting moiety is attached to a liposome. For example, US
5 Patent 6,245,427 describes a method for targeting a liposome using a protein or peptide. In
another example, a cationic lipid component of the liposome is derivatized with a targeting
moiety. For example, WO 96/37194 describes converting N-glutaryl dioleoylphosphatidyl
ethanolamine to a N-hydroxysuccinimide activated ester. The product was then coupled to an
RGD peptide.

10 GENES AND DISEASES

In one aspect, the invention features, a method of treating a subject at risk for or afflicted
with unwanted cell proliferation, *e.g.*, malignant or nonmalignant cell proliferation. The method
includes:

15 providing an iRNA agent, *e.g.*, an sRNA or iRNA agent described herein, *e.g.*, an iRNA
having a structure described herein, where the iRNA is homologous to and can silence, *e.g.*, by
cleavage, a gene which promotes unwanted cell proliferation;

administering an iRNA agent, *e.g.*, an sRNA or iRNA agent described herein to a subject,
preferably a human subject,

thereby treating the subject.

20 In a preferred embodiment the gene is a growth factor or growth factor receptor gene, a
kinase, *e.g.*, a protein tyrosine, serine or threonine kinase gene, an adaptor protein gene, a gene
encoding a G protein superfamily molecule, or a gene encoding a transcription factor.

In a preferred embodiment the iRNA agent silences the PDGF beta gene, and thus can be
used to treat a subject having or at risk for a disorder characterized by unwanted PDGF beta
25 expression, *e.g.*, testicular and lung cancers.

Attorney's Docket No.: 14174-072W01

In another preferred embodiment the iRNA agent silences the Erb-B gene, and thus can be used to treat a subject having or at risk for a disorder characterized by unwanted Erb-B expression, *e.g.*, breast cancer.

- 5 In a preferred embodiment the iRNA agent silences the Src gene, and thus can be used to treat a subject having or at risk for a disorder characterized by unwanted Src expression, *e.g.*, colon cancers.

In a preferred embodiment the iRNA agent silences the CRK gene, and thus can be used to treat a subject having or at risk for a disorder characterized by unwanted CRK expression, *e.g.*, colon and lung cancers.

- 10 In a preferred embodiment the iRNA agent silences the GRB2 gene, and thus can be used to treat a subject having or at risk for a disorder characterized by unwanted GRB2 expression, *e.g.*, squamous cell carcinoma.

- In another preferred embodiment the iRNA agent silences the RAS gene, and thus can be used to treat a subject having or at risk for a disorder characterized by unwanted RAS
15 expression, *e.g.*, pancreatic, colon and lung cancers, and chronic leukemia.

In another preferred embodiment the iRNA agent silences the MEKK gene, and thus can be used to treat a subject having or at risk for a disorder characterized by unwanted MEKK expression, *e.g.*, squamous cell carcinoma, melanoma or leukemia.

- In another preferred embodiment the iRNA agent silences the JNK gene, and thus can be
20 used to treat a subject having or at risk for a disorder characterized by unwanted JNK expression, *e.g.*, pancreatic or breast cancers.

In a preferred embodiment the iRNA agent silences the RAF gene, and thus can be used to treat a subject having or at risk for a disorder characterized by unwanted RAF expression, *e.g.*, lung cancer or leukemia.

- 25 In a preferred embodiment the iRNA agent silences the Erk1/2 gene, and thus can be used to treat a subject having or at risk for a disorder characterized by unwanted Erk1/2 expression, *e.g.*, lung cancer.

Attorney's Docket No.: 14174-072W01

In another preferred embodiment the iRNA agent silences the PCNA(p21) gene, and thus can be used to treat a subject having or at risk for a disorder characterized by unwanted PCNA expression, *e.g.*, lung cancer.

5 In a preferred embodiment the iRNA agent silences the MYB gene, and thus can be used to treat a subject having or at risk for a disorder characterized by unwanted MYB expression, *e.g.*, colon cancer or chronic myelogenous leukemia.

In a preferred embodiment the iRNA agent silences the c-MYC gene, and thus can be used to treat a subject having or at risk for a disorder characterized by unwanted c-MYC expression, *e.g.*, Burkitt's Lymphoma or neuroblastoma.

10 In another preferred embodiment the iRNA agent silences the JUN gene, and thus can be used to treat a subject having or at risk for a disorder characterized by unwanted JUN expression, *e.g.*, ovarian, prostate or breast cancers.

In another preferred embodiment the iRNA agent silences the FOS gene, and thus can be used to treat a subject having or at risk for a disorder characterized by unwanted FOS expression,
15 *e.g.*, skin or prostate cancers.

In a preferred embodiment the iRNA agent silences the BCL-2 gene, and thus can be used to treat a subject having or at risk for a disorder characterized by unwanted BCL-2 expression, *e.g.*, lung or prostate cancers or Non-Hodgkin lymphoma.

20 In a preferred embodiment the iRNA agent silences the Cyclin D gene, and thus can be used to treat a subject having or at risk for a disorder characterized by unwanted Cyclin D expression, *e.g.*, esophageal and colon cancers.

In a preferred embodiment the iRNA agent silences the VEGF gene, and thus can be used to treat a subject having or at risk for a disorder characterized by unwanted VEGF expression, *e.g.*, esophageal and colon cancers.

25 In a preferred embodiment the iRNA agent silences the EGFR gene, and thus can be used to treat a subject having or at risk for a disorder characterized by unwanted EGFR expression, *e.g.*, breast cancer.

Attorney's Docket No.: 14174-072W01

In another preferred embodiment the iRNA agent silences the Cyclin A gene, and thus can be used to treat a subject having or at risk for a disorder characterized by unwanted Cyclin A expression, *e.g.*, lung and cervical cancers.

5 In another preferred embodiment the iRNA agent silences the Cyclin E gene, and thus can be used to treat a subject having or at risk for a disorder characterized by unwanted Cyclin E expression, *e.g.*, lung and breast cancers.

In another preferred embodiment the iRNA agent silences the WNT-1 gene, and thus can be used to treat a subject having or at risk for a disorder characterized by unwanted WNT-1 expression, *e.g.*, basal cell carcinoma.

10 In another preferred embodiment the iRNA agent silences the beta-catenin gene, and thus can be used to treat a subject having or at risk for a disorder characterized by unwanted beta-catenin expression, *e.g.*, adenocarcinoma or hepatocellular carcinoma.

In another preferred embodiment the iRNA agent silences the c-MET gene, and thus can be used to treat a subject having or at risk for a disorder characterized by unwanted c-MET
15 expression, *e.g.*, hepatocellular carcinoma.

In another preferred embodiment the iRNA agent silences the PKC gene, and thus can be used to treat a subject having or at risk for a disorder characterized by unwanted PKC expression, *e.g.*, breast cancer.

In a preferred embodiment the iRNA agent silences the NFkB gene, and thus can be used
20 to treat a subject having or at risk for a disorder characterized by unwanted NFkB expression, *e.g.*, breast cancer.

In a preferred embodiment the iRNA agent silences the STAT3 gene, and thus can be used to treat a subject having or at risk for a disorder characterized by unwanted STAT3 expression, *e.g.*, prostate cancer.

25 In another preferred embodiment the iRNA agent silences the survivin gene, and thus can be used to treat a subject having or at risk for a disorder characterized by unwanted survivin expression, *e.g.*, cervical or pancreatic cancers.

Attorney's Docket No.: 14174-072W01

In another preferred embodiment the iRNA agent silences the Her2/Neu gene, and thus can be used to treat a subject having or at risk for a disorder characterized by unwanted Her2/Neu expression, *e.g.*, breast cancer.

5 In another preferred embodiment the iRNA agent silences the topoisomerase I gene, and thus can be used to treat a subject having or at risk for a disorder characterized by unwanted topoisomerase I expression, *e.g.*, ovarian and colon cancers.

In a preferred embodiment the iRNA agent silences the topoisomerase II alpha gene, and thus can be used to treat a subject having or at risk for a disorder characterized by unwanted topoisomerase II expression, *e.g.*, breast and colon cancers.

10 In a preferred embodiment the iRNA agent silences mutations in the p73 gene, and thus can be used to treat a subject having or at risk for a disorder characterized by unwanted p73 expression, *e.g.*, colorectal adenocarcinoma.

In a preferred embodiment the iRNA agent silences mutations in the p21(WAF1/CIP1) gene, and thus can be used to treat a subject having or at risk for a disorder characterized by
15 unwanted p21(WAF1/CIP1) expression, *e.g.*, liver cancer.

In a preferred embodiment the iRNA agent silences mutations in the p27(KIP1) gene, and thus can be used to treat a subject having or at risk for a disorder characterized by unwanted p27(KIP1) expression, *e.g.*, liver cancer.

In preferred embodiments the iRNA agent silences mutations in tumor suppressor genes,
20 and thus can be used as a method to promote apoptotic activity in combination with chemotherapeutics.

In another aspect, the invention features, a method of treating a subject, *e.g.*, a human, at risk for or afflicted with a disease or disorder that may benefit by angiogenesis inhibition *e.g.*, cancer. The method includes:

25 providing an iRNA agent, *e.g.*, an iRNA agent having a structure described herein, which iRNA agent is homologous to and can silence, *e.g.*, by cleavage, a gene which mediates angiogenesis;

Attorney's Docket No.: 14174-072W01

administering the iRNA agent to a subject,
thereby treating the subject.

In another aspect, the invention features a method of treating a subject infected with a virus or at risk for or afflicted with a disorder or disease associated with a viral infection. The
5 method includes:

providing an iRNA agent, *e.g.*, and iRNA agent having a structure described herein,
which iRNA agent is homologous to and can silence, *e.g.*, by cleavage, a viral gene of a cellular
gene which mediates viral function, *e.g.*, entry or growth;

administering the iRNA agent to a subject, preferably a human subject,
10 thereby treating the subject.

Thus, the invention provides for a method of treating patients infected by the Human Papilloma Virus (HPV) or at risk for or afflicted with a disorder mediated by HPV, *e.g.*, cervical cancer. HPV is linked to 95% of cervical carcinomas and thus an antiviral therapy is an attractive method to treat these cancers and other symptoms of viral infection.

15 In a preferred embodiment, the expression of a HPV gene is reduced. In another preferred embodiment, the HPV gene is one of the group of E2, E6, or E7.

In a preferred embodiment the expression of a human gene that is required for HPV replication is reduced.

The invention also includes a method of treating patients infected by the Human
20 Immunodeficiency Virus (HIV) or at risk for or afflicted with a disorder mediated by HIV, *e.g.*, Acquired Immune Deficiency Syndrome (AIDS).

In a preferred embodiment, the expression of a HIV gene is reduced. In another preferred embodiment, the HIV gene is CCR5, Gag, or Rev.

In a preferred embodiment the expression of a human gene that is required for HIV
25 replication is reduced. In another preferred embodiment, the gene is CD4 or Tsg101.

Attorney's Docket No.: 14174-072W01

The invention also includes a method for treating patients infected by the Hepatitis B Virus (HBV) or at risk for or afflicted with a disorder mediated by HBV, *e.g.*, cirrhosis and hepatocellular carcinoma.

5 In a preferred embodiment, the expression of a HBV gene is reduced. In another preferred embodiment, the targeted HBV gene encodes one of the group of the tail region of the HBV core protein, the pre-c region, or the c region. In another preferred embodiment, a targeted HBV-RNA sequence is comprised of the poly(A) tail.

In preferred embodiment the expression of a human gene that is required for HBV replication is reduced.

10 The invention also provides for a method of treating patients infected by the Hepatitis A Virus (HAV), or at risk for or afflicted with a disorder mediated by HAV.

In a preferred embodiment the expression of a human gene that is required for HAV replication is reduced.

15 The present invention provides for a method of treating patients infected by the Hepatitis C Virus (HCV), or at risk for or afflicted with a disorder mediated by HCV, *e.g.*, cirrhosis

In a preferred embodiment, the expression of a HCV gene is reduced.

In another preferred embodiment the expression of a human gene that is required for HCV replication is reduced.

20 The present invention also provides for a method of treating patients infected by the any of the group of Hepatitis Viral strains comprising hepatitis D, E, F, G, or H, or patients at risk for or afflicted with a disorder mediated by any of these strains of hepatitis.

In a preferred embodiment, the expression of a Hepatitis, D, E, F, G, or H gene is reduced.

25 In another preferred embodiment the expression of a human gene that is required for hepatitis D, E, F, G or H replication is reduced.

Attorney's Docket No.: 14174-072W01

Methods of the invention also provide for treating patients infected by the Respiratory Syncytial Virus (RSV) or at risk for or afflicted with a disorder mediated by RSV, *e.g.*, lower respiratory tract infection in infants and childhood asthma, pneumonia and other complications, *e.g.*, in the elderly.

5 In a preferred embodiment, the expression of a RSV gene is reduced. In another preferred embodiment, the targeted HBV gene encodes one of the group of genes N, L, or P.

In a preferred embodiment the expression of a human gene that is required for RSV replication is reduced.

10 Methods of the invention provide for treating patients infected by the Herpes Simplex Virus (HSV) or at risk for or afflicted with a disorder mediated by HSV, *e.g.*, genital herpes and cold sores as well as life-threatening or sight-impairing disease mainly in immunocompromised patients.

In a preferred embodiment, the expression of a HSV gene is reduced. In another preferred embodiment, the targeted HSV gene encodes DNA polymerase or the helicase-primase.
15

In a preferred embodiment the expression of a human gene that is required for HSV replication is reduced.

The invention also provides a method for treating patients infected by the herpes Cytomegalovirus (CMV) or at risk for or afflicted with a disorder mediated by CMV, *e.g.*,
20 congenital virus infections and morbidity in immunocompromised patients.

In a preferred embodiment, the expression of a CMV gene is reduced.

In a preferred embodiment the expression of a human gene that is required for CMV replication is reduced.

25 Methods of the invention also provide for a method of treating patients infected by the herpes Epstein Barr Virus (EBV) or at risk for or afflicted with a disorder mediated by EBV, *e.g.*, NK/T-cell lymphoma, non-Hodgkin lymphoma, and Hodgkin disease.

Attorney's Docket No.: 14174-072W01

In a preferred embodiment, the expression of a EBV gene is reduced.

In a preferred embodiment the expression of a human gene that is required for EBV replication is reduced.

5 Methods of the invention also provide for treating patients infected by Kaposi's Sarcoma-associated Herpes Virus (KSHV), also called human herpesvirus 8, or patients at risk for or afflicted with a disorder mediated by KSHV, *e.g.*, Kaposi's sarcoma, multicentric Castleman's disease and AIDS-associated primary effusion lymphoma.

In a preferred embodiment, the expression of a KSHV gene is reduced.

10 In a preferred embodiment the expression of a human gene that is required for KSHV replication is reduced.

The invention also includes a method for treating patients infected by the JC Virus (JCV) or a disease or disorder associated with this virus, *e.g.*, progressive multifocal leukoencephalopathy (PML).

In a preferred embodiment, the expression of a JCV gene is reduced.

15 In preferred embodiment the expression of a human gene that is required for JCV replication is reduced.

Methods of the invention also provide for treating patients infected by the myxovirus or at risk for or afflicted with a disorder mediated by myxovirus, *e.g.*, influenza.

In a preferred embodiment, the expression of a myxovirus gene is reduced.

20 In a preferred embodiment the expression of a human gene that is required for myxovirus replication is reduced.

Methods of the invention also provide for treating patients infected by the rhinovirus or at risk for or afflicted with a disorder mediated by rhinovirus, *e.g.*, the common cold.

In a preferred embodiment, the expression of a rhinovirus gene is reduced.

Attorney's Docket No.: 14174-072W01

In preferred embodiment the expression of a human gene that is required for rhinovirus replication is reduced.

Methods of the invention also provide for treating patients infected by the coronavirus or at risk for or afflicted with a disorder mediated by coronavirus, *e.g.*, the common cold.

5 In a preferred embodiment, the expression of a coronavirus gene is reduced.

In preferred embodiment the expression of a human gene that is required for coronavirus replication is reduced.

Methods of the invention also provide for treating patients infected by the flavivirus West Nile or at risk for or afflicted with a disorder mediated by West Nile Virus.

10 In a preferred embodiment, the expression of a West Nile Virus gene is reduced. In another preferred embodiment, the West Nile Virus gene is one of the group comprising E, NS3, or NS5.

In a preferred embodiment the expression of a human gene that is required for West Nile Virus replication is reduced.

15 Methods of the invention also provide for treating patients infected by the St. Louis Encephalitis flavivirus, or at risk for or afflicted with a disease or disorder associated with this virus, *e.g.*, viral haemorrhagic fever or neurological disease.

In a preferred embodiment, the expression of a St. Louis Encephalitis gene is reduced.

20 In a preferred embodiment the expression of a human gene that is required for St. Louis Encephalitis virus replication is reduced.

Methods of the invention also provide for treating patients infected by the Tick-borne encephalitis flavivirus, or at risk for or afflicted with a disorder mediated by Tick-borne encephalitis virus, *e.g.*, viral haemorrhagic fever and neurological disease.

25 In a preferred embodiment, the expression of a Tick-borne encephalitis virus gene is reduced.

Attorney's Docket No.: 14174-072W01

In a preferred embodiment the expression of a human gene that is required for Tick-borne encephalitis virus replication is reduced.

Methods of the invention also provide for methods of treating patients infected by the Murray Valley encephalitis flavivirus, which commonly results in viral haemorrhagic fever and
5 neurological disease.

In a preferred embodiment, the expression of a Murray Valley encephalitis virus gene is reduced.

In a preferred embodiment the expression of a human gene that is required for Murray Valley encephalitis virus replication is reduced.

10 The invention also includes methods for treating patients infected by the dengue flavivirus, or a disease or disorder associated with this virus, *e.g.*, dengue haemorrhagic fever.

In a preferred embodiment, the expression of a dengue virus gene is reduced.

In a preferred embodiment the expression of a human gene that is required for dengue virus replication is reduced.

15 Methods of the invention also provide for treating patients infected by the Simian Virus 40 (SV40) or at risk for or afflicted with a disorder mediated by SV40, *e.g.*, tumorigenesis.

In a preferred embodiment, the expression of a SV40 gene is reduced.

In a preferred embodiment the expression of a human gene that is required for SV40 replication is reduced.

20 The invention also includes methods for treating patients infected by the Human T Cell Lymphotropic Virus (HTLV), or a disease or disorder associated with this virus, *e.g.*, leukemia and myelopathy.

In a preferred embodiment, the expression of a HTLV gene is reduced. In another preferred embodiment the HTLV1 gene is the Tax transcriptional activator.

Attorney's Docket No.: 14174-072W01

In a preferred embodiment the expression of a human gene that is required for HTLV replication is reduced.

Methods of the invention also provide for treating patients infected by the Moloney-Murine Leukemia Virus (Mo-MuLV) or at risk for or afflicted with a disorder mediated by Mo-MuLV, *e.g.*, T-cell leukemia.

In a preferred embodiment, the expression of a Mo-MuLV gene is reduced.

In a preferred embodiment the expression of a human gene that is required for Mo-MuLV replication is reduced.

Methods of the invention also provide for treating patients infected by the encephalomyocarditis virus (EMCV) or at risk for or afflicted with a disorder mediated by EMCV, *e.g.* myocarditis. EMCV leads to myocarditis in mice and pigs and is capable of infecting human myocardial cells. This virus is therefore a concern for patients undergoing xenotransplantation.

In a preferred embodiment, the expression of a EMCV gene is reduced.

In a preferred embodiment the expression of a human gene that is required for EMCV replication is reduced.

The invention also includes a method for treating patients infected by the measles virus (MV) or at risk for or afflicted with a disorder mediated by MV, *e.g.* measles.

In a preferred embodiment, the expression of a MV gene is reduced.

In a preferred embodiment the expression of a human gene that is required for MV replication is reduced.

The invention also includes a method for treating patients infected by the Varicella zoster virus (VZV) or at risk for or afflicted with a disorder mediated by VZV, *e.g.* chicken pox or shingles (also called zoster).

In a preferred embodiment, the expression of a VZV gene is reduced.

Attorney's Docket No.: 14174-072W01

In a preferred embodiment the expression of a human gene that is required for VZV replication is reduced.

The invention also includes a method for treating patients infected by an adenovirus or at risk for or afflicted with a disorder mediated by an adenovirus, *e.g.* respiratory tract infection.

5 In a preferred embodiment, the expression of an adenovirus gene is reduced.

In a preferred embodiment the expression of a human gene that is required for adenovirus replication is reduced.

The invention includes a method for treating patients infected by a yellow fever virus (YFV) or at risk for or afflicted with a disorder mediated by a YFV, *e.g.* respiratory tract
10 infection.

In a preferred embodiment, the expression of a YFV gene is reduced. In another preferred embodiment, the preferred gene is one of a group that includes the E, NS2A, or NS3 genes.

In a preferred embodiment the expression of a human gene that is required for YFV
15 replication is reduced.

Methods of the invention also provide for treating patients infected by the poliovirus or at risk for or afflicted with a disorder mediated by poliovirus, *e.g.*, polio.

In a preferred embodiment, the expression of a poliovirus gene is reduced.

In a preferred embodiment the expression of a human gene that is required for poliovirus
20 replication is reduced.

Methods of the invention also provide for treating patients infected by a poxvirus or at risk for or afflicted with a disorder mediated by a poxvirus, *e.g.*, smallpox

In a preferred embodiment, the expression of a poxvirus gene is reduced.

Attorney's Docket No.: 14174-072W01

In a preferred embodiment the expression of a human gene that is required for poxvirus replication is reduced.

In another, aspect the invention features methods of treating a subject infected with a pathogen, *e.g.*, a bacterial, amoebic, parasitic, or fungal pathogen. The method includes:

- 5 providing a iRNA agent, *e.g.*, a siRNA having a structure described herein, where siRNA is homologous to and can silence, *e.g.*, by cleavage of a pathogen gene;

administering the iRNA agent to a subject, preferably a human subject,

thereby treating the subject.

- The target gene can be one involved in growth, cell wall synthesis, protein synthesis,
10 transcription, energy metabolism, *e.g.*, the Krebs cycle, or toxin production.

Thus, the present invention provides for a method of treating patients infected by a plasmodium that causes malaria.

In a preferred embodiment, the expression of a plasmodium gene is reduced. In another preferred embodiment, the gene is apical membrane antigen 1 (AMA1).

- 15 In a preferred embodiment the expression of a human gene that is required for plasmodium replication is reduced.

The invention also includes methods for treating patients infected by the Mycobacterium ulcerans, or a disease or disorder associated with this pathogen, *e.g.* Buruli ulcers.

In a preferred embodiment, the expression of a Mycobacterium ulcerans gene is reduced.

- 20 In a preferred embodiment the expression of a human gene that is required for Mycobacterium ulcerans replication is reduced.

The invention also includes methods for treating patients infected by the Mycobacterium tuberculosis, or a disease or disorder associated with this pathogen, *e.g.* tuberculosis.

Attorney's Docket No.: 14174-072W01

In a preferred embodiment, the expression of a *Mycobacterium tuberculosis* gene is reduced.

In a preferred embodiment the expression of a human gene that is required for *Mycobacterium tuberculosis* replication is reduced.

- 5 The invention also includes methods for treating patients infected by the *Mycobacterium leprae*, or a disease or disorder associated with this pathogen, *e.g.* leprosy.

In a preferred embodiment, the expression of a *Mycobacterium leprae* gene is reduced.

In a preferred embodiment the expression of a human gene that is required for *Mycobacterium leprae* replication is reduced.

- 10 The invention also includes methods for treating patients infected by the bacteria *Staphylococcus aureus*, or a disease or disorder associated with this pathogen, *e.g.* infections of the skin and mucous membranes.

In a preferred embodiment, the expression of a *Staphylococcus aureus* gene is reduced.

- 15 In a preferred embodiment the expression of a human gene that is required for *Staphylococcus aureus* replication is reduced.

The invention also includes methods for treating patients infected by the bacteria *Streptococcus pneumoniae*, or a disease or disorder associated with this pathogen, *e.g.* pneumonia or childhood lower respiratory tract infection.

- 20 In a preferred embodiment, the expression of a *Streptococcus pneumoniae* gene is reduced.

In a preferred embodiment the expression of a human gene that is required for *Streptococcus pneumoniae* replication is reduced.

- 25 The invention also includes methods for treating patients infected by the bacteria *Streptococcus pyogenes*, or a disease or disorder associated with this pathogen, *e.g.* Strep throat or Scarlet fever.

Attorney's Docket No.: 14174-072W01

In a preferred embodiment, the expression of a *Streptococcus pyogenes* gene is reduced.

In a preferred embodiment the expression of a human gene that is required for *Streptococcus pyogenes* replication is reduced.

- The invention also includes methods for treating patients infected by the bacteria
- 5 Chlamydia pneumoniae, or a disease or disorder associated with this pathogen, *e.g.* pneumonia or childhood lower respiratory tract infection

In a preferred embodiment, the expression of a *Chlamydia pneumoniae* gene is reduced.

In a preferred embodiment the expression of a human gene that is required for *Chlamydia pneumoniae* replication is reduced.

- The invention also includes methods for treating patients infected by the bacteria
- 10 Mycoplasma pneumoniae, or a disease or disorder associated with this pathogen, *e.g.* pneumonia or childhood lower respiratory tract infection

In a preferred embodiment, the expression of a *Mycoplasma pneumoniae* gene is reduced.

- In a preferred embodiment the expression of a human gene that is required for
- 15 Mycoplasma pneumoniae replication is reduced.

- The loss of heterozygosity (LOH) can result in hemizygosity for sequence, *e.g.*, genes, in the area of LOH. This can result in a significant genetic difference between normal and disease-state cells, *e.g.*, cancer cells, and provides a useful difference between normal and disease-state cells, *e.g.*, cancer cells. This difference can arise because a gene or other sequence is
- 20 heterozygous in euploid cells but is hemizygous in cells having LOH. The regions of LOH will often include a gene, the loss of which promotes unwanted proliferation, *e.g.*, a tumor suppressor gene, and other sequences including, *e.g.*, other genes, in some cases a gene which is essential for normal function, *e.g.*, growth. Methods of the invention rely, in part, on the specific cleavage or silencing of one allele of an essential gene with an iRNA agent of the invention. The iRNA
- 25 agent is selected such that it targets the single allele of the essential gene found in the cells having LOH but does not silence the other allele, which is present in cells which do not show LOH. In essence, it discriminates between the two alleles, preferentially silencing the selected

Attorney's Docket No.: 14174-072W01

allele. In essence polymorphisms, *e.g.*, SNPs of essential genes that are affected by LOH, are used as a target for a disorder characterized by cells having LOH, *e.g.*, cancer cells having LOH.

E.g., one of ordinary skill in the art can identify essential genes which are in proximity to tumor suppressor genes, and which are within a LOH region which includes the tumor
5 suppressor gene. The gene encoding the large subunit of human RNA polymerase II, POLR2A, a gene located in close proximity to the tumor suppressor gene p53, is such a gene. It frequently occurs within a region of LOH in cancer cells. Other genes that occur within LOH regions and are lost in many cancer cell types include the group comprising replication protein A 70-kDa subunit, replication protein A 32-kD, ribonucleotide reductase, thymidilate synthase, TATA
10 associated factor 2H, ribosomal protein S14, eukaryotic initiation factor 5A, alanyl tRNA synthetase, cysteinyl tRNA synthetase, NaK ATPase, alpha-1 subunit, and transferrin receptor.

Accordingly, the invention features, a method of treating a disorder characterized by LOH, *e.g.*, cancer. The method includes:

optionally, determining the genotype of the allele of a gene in the region of LOH and
15 preferably determining the genotype of both alleles of the gene in a normal cell;

providing an iRNA agent which preferentially cleaves or silences the allele found in the LOH cells;

administering the iRNA to the subject,

thereby treating the disorder.

20 The invention also includes a iRNA agent disclosed herein, *e.g.*, an iRNA agent which can preferentially silence, *e.g.*, cleave, one allele of a polymorphic gene

In another aspect, the invention provides a method of cleaving or silencing more than one gene with an iRNA agent. In these embodiments the iRNA agent is selected so that it has sufficient homology to a sequence found in more than one gene. For example, the sequence
25 AAGCTGGCCCTGGACATGGAGAT (SEQ ID NO:6719) is conserved between mouse lamin B1, lamin B2, keratin complex 2-gene 1 and lamin A/C. Thus an iRNA agent targeted to this sequence would effectively silence the entire collection of genes.

Attorney's Docket No.: 14174-072W01

The invention also includes an iRNA agent disclosed herein, which can silence more than one gene.

ROUTE OF DELIVERY

For ease of exposition the formulations, compositions and methods in this section are discussed largely with regard to unmodified iRNA agents. It should be understood, however, that these formulations, compositions and methods can be practiced with other iRNA agents, *e.g.*, modified iRNA agents, and such practice is within the invention. A composition that includes a iRNA can be delivered to a subject by a variety of routes. Exemplary routes include: intravenous, topical, rectal, anal, vaginal, nasal, pulmonary, ocular.

The iRNA molecules of the invention can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically include one or more species of iRNA and a pharmaceutically acceptable carrier. As used herein the language "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

The pharmaceutical compositions of the present invention may be administered in a number of ways depending upon whether local or systemic treatment is desired and upon the area to be treated. Administration may be topical (including ophthalmic, vaginal, rectal, intranasal, transdermal), oral or parenteral. Parenteral administration includes intravenous drip, subcutaneous, intraperitoneal or intramuscular injection, or intrathecal or intraventricular administration.

The route and site of administration may be chosen to enhance targeting. For example, to target muscle cells, intramuscular injection into the muscles of interest would be a logical choice. Lung cells might be targeted by administering the iRNA in aerosol form. The vascular

Attorney's Docket No.: 14174-072W01

endothelial cells could be targeted by coating a balloon catheter with the iRNA and mechanically introducing the DNA.

Formulations for topical administration may include transdermal patches, ointments, lotions, creams, gels, drops, suppositories, sprays, liquids and powders. Conventional pharmaceutical carriers, aqueous, powder or oily bases, thickeners and the like may be necessary or desirable. Coated condoms, gloves and the like may also be useful.

Compositions for oral administration include powders or granules, suspensions or solutions in water, syrups, elixirs or non-aqueous media, tablets, capsules, lozenges, or troches. In the case of tablets, carriers that can be used include lactose, sodium citrate and salts of phosphoric acid. Various disintegrants such as starch, and lubricating agents such as magnesium stearate, sodium lauryl sulfate and talc, are commonly used in tablets. For oral administration in capsule form, useful diluents are lactose and high molecular weight polyethylene glycols. When aqueous suspensions are required for oral use, the nucleic acid compositions can be combined with emulsifying and suspending agents. If desired, certain sweetening and/or flavoring agents can be added.

Compositions for intrathecal or intraventricular administration may include sterile aqueous solutions which may also contain buffers, diluents and other suitable additives.

Formulations for parenteral administration may include sterile aqueous solutions which may also contain buffers, diluents and other suitable additives. Intraventricular injection may be facilitated by an intraventricular catheter, for example, attached to a reservoir. For intravenous use, the total concentration of solutes should be controlled to render the preparation isotonic.

For ocular administration, ointments or droppable liquids may be delivered by ocular delivery systems known to the art such as applicators or eye droppers. Such compositions can include mucomimetics such as hyaluronic acid, chondroitin sulfate, hydroxypropyl methylcellulose or poly(vinyl alcohol), preservatives such as sorbic acid, EDTA or benzylchromium chloride, and the usual quantities of diluents and/or carriers.

Topical Delivery

For ease of exposition the formulations, compositions and methods in this section are discussed largely with regard to unmodified iRNA agents. It should be understood, however, that these formulations, compositions and methods can be practiced with other iRNA agents, *e.g.*, modified iRNA agents, and such practice is within the invention. In a preferred embodiment, an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, (*e.g.*, a precursor, *e.g.*, a larger iRNA agent which can be processed into a sRNA agent, or a DNA which encodes an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, or precursor thereof) is delivered to a subject via topical administration. "Topical administration" refers to the delivery to a subject by contacting the formulation directly to a surface of the subject. The most common form of topical delivery is to the skin, but a composition disclosed herein can also be directly applied to other surfaces of the body, *e.g.*, to the eye, a mucous membrane, to surfaces of a body cavity or to an internal surface. As mentioned above, the most common topical delivery is to the skin. The term encompasses several routes of administration including, but not limited to, topical and transdermal. These modes of administration typically include penetration of the skin's permeability barrier and efficient delivery to the target tissue or stratum. Topical administration can be used as a means to penetrate the epidermis and dermis and ultimately achieve systemic delivery of the composition. Topical administration can also be used as a means to selectively deliver oligonucleotides to the epidermis or dermis of a subject, or to specific strata thereof, or to an underlying tissue.

The term "skin," as used herein, refers to the epidermis and/or dermis of an animal. Mammalian skin consists of two major, distinct layers. The outer layer of the skin is called the epidermis. The epidermis is comprised of the stratum corneum, the stratum granulosum, the stratum spinosum, and the stratum basale, with the stratum corneum being at the surface of the skin and the stratum basale being the deepest portion of the epidermis. The epidermis is between 50 μ m and 0.2 mm thick, depending on its location on the body.

Beneath the epidermis is the dermis, which is significantly thicker than the epidermis. The dermis is primarily composed of collagen in the form of fibrous bundles. The collagenous

bundles provide support for, inter alia, blood vessels, lymph capillaries, glands, nerve endings and immunologically active cells.

One of the major functions of the skin as an organ is to regulate the entry of substances into the body. The principal permeability barrier of the skin is provided by the stratum corneum, which is formed from many layers of cells in various states of differentiation. The spaces between cells in the stratum corneum is filled with different lipids arranged in lattice-like formations that provide seals to further enhance the skins permeability barrier.

The permeability barrier provided by the skin is such that it is largely impermeable to molecules having molecular weight greater than about 750 Da. For larger molecules to cross the skin's permeability barrier, mechanisms other than normal osmosis must be used.

Several factors determine the permeability of the skin to administered agents. These factors include the characteristics of the treated skin, the characteristics of the delivery agent, interactions between both the drug and delivery agent and the drug and skin, the dosage of the drug applied, the form of treatment, and the post treatment regimen. To selectively target the epidermis and dermis, it is sometimes possible to formulate a composition that comprises one or more penetration enhancers that will enable penetration of the drug to a preselected stratum.

Transdermal delivery is a valuable route for the administration of lipid soluble therapeutics. The dermis is more permeable than the epidermis and therefore absorption is much more rapid through abraded, burned or denuded skin. Inflammation and other physiologic conditions that increase blood flow to the skin also enhance transdermal adsorption. Absorption via this route may be enhanced by the use of an oily vehicle (inunction) or through the use of one or more penetration enhancers. Other effective ways to deliver a composition disclosed herein via the transdermal route include hydration of the skin and the use of controlled release topical patches. The transdermal route provides a potentially effective means to deliver a composition disclosed herein for systemic and/or local therapy.

In addition, iontophoresis (transfer of ionic solutes through biological membranes under the influence of an electric field) (Lee *et al.*, Critical Reviews in Therapeutic Drug Carrier Systems, 1991, p. 163), phonophoresis or sonophoresis (use of ultrasound to enhance the

Attorney's Docket No.: 14174-072W01

absorption of various therapeutic agents across biological membranes, notably the skin and the cornea) (Lee *et al.*, Critical Reviews in Therapeutic Drug Carrier Systems, 1991, p. 166), and optimization of vehicle characteristics relative to dose position and retention at the site of administration (Lee *et al.*, Critical Reviews in Therapeutic Drug Carrier Systems, 1991, p. 168) may be useful methods for enhancing the transport of topically applied compositions across skin and mucosal sites.

The compositions and methods provided may also be used to examine the function of various proteins and genes *in vitro* in cultured or preserved dermal tissues and in animals. The invention can be thus applied to examine the function of any gene. The methods of the invention can also be used therapeutically or prophylactically. For example, for the treatment of animals that are known or suspected to suffer from diseases such as psoriasis, lichen planus, toxic epidermal necrolysis, erythema multiforme, basal cell carcinoma, squamous cell carcinoma, malignant melanoma, Paget's disease, Kaposi's sarcoma, pulmonary fibrosis, Lyme disease and viral, fungal and bacterial infections of the skin.

Pulmonary Delivery

For ease of exposition the formulations, compositions and methods in this section are discussed largely with regard to unmodified iRNA agents. It should be understood, however, that these formulations, compositions and methods can be practiced with other iRNA agents, *e.g.*, modified iRNA agents, and such practice is within the invention. A composition that includes an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, (*e.g.*, a precursor, *e.g.*, a larger iRNA agent which can be processed into a sRNA agent, or a DNA which encodes an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, or precursor thereof) can be administered to a subject by pulmonary delivery. Pulmonary delivery compositions can be delivered by inhalation by the patient of a dispersion so that the composition, preferably iRNA, within the dispersion can reach the lung where it can be readily absorbed through the alveolar region directly into blood circulation. Pulmonary delivery can be effective both for systemic delivery and for localized delivery to treat diseases of the lungs.

Pulmonary delivery can be achieved by different approaches, including the use of nebulized, aerosolized, micellar and dry powder-based formulations. Delivery can be achieved

with liquid nebulizers, aerosol-based inhalers, and dry powder dispersion devices. Metered-dose devices are preferred. One of the benefits of using an atomizer or inhaler is that the potential for contamination is minimized because the devices are self contained. Dry powder dispersion devices, for example, deliver drugs that may be readily formulated as dry powders. A iRNA composition may be stably stored as lyophilized or spray-dried powders by itself or in combination with suitable powder carriers. The delivery of a composition for inhalation can be mediated by a dosing timing element which can include a timer, a dose counter, time measuring device, or a time indicator which when incorporated into the device enables dose tracking, compliance monitoring, and/or dose triggering to a patient during administration of the aerosol medicament.

The term "powder" means a composition that consists of finely dispersed solid particles that are free flowing and capable of being readily dispersed in an inhalation device and subsequently inhaled by a subject so that the particles reach the lungs to permit penetration into the alveoli. Thus, the powder is said to be "respirable." Preferably the average particle size is less than about 10 μm in diameter preferably with a relatively uniform spheroidal shape distribution. More preferably the diameter is less than about 7.5 μm and most preferably less than about 5.0 μm . Usually the particle size distribution is between about 0.1 μm and about 5 μm in diameter, particularly about 0.3 μm to about 5 μm .

The term "dry" means that the composition has a moisture content below about 10% by weight (% w) water, usually below about 5% w and preferably less than about 3% w. A dry composition can be such that the particles are readily dispersible in an inhalation device to form an aerosol.

The term "therapeutically effective amount" is the amount present in the composition that is needed to provide the desired level of drug in the subject to be treated to give the anticipated physiological response.

The term "physiologically effective amount" is that amount delivered to a subject to give the desired palliative or curative effect.

The term "pharmaceutically acceptable carrier" means that the carrier can be taken into the lungs with no significant adverse toxicological effects on the lungs.

The types of pharmaceutical excipients that are useful as carrier include stabilizers such as human serum albumin (HSA), bulking agents such as carbohydrates, amino acids and polypeptides; pH adjusters or buffers; salts such as sodium chloride; and the like. These carriers may be in a crystalline or amorphous form or may be a mixture of the two.

Bulking agents that are particularly valuable include compatible carbohydrates, polypeptides, amino acids or combinations thereof. Suitable carbohydrates include monosaccharides such as galactose, D-mannose, sorbose, and the like; disaccharides, such as lactose, trehalose, and the like; cyclodextrins, such as 2-hydroxypropyl- β -cyclodextrin; and polysaccharides, such as raffinose, maltodextrins, dextrans, and the like; alditols, such as mannitol, xylitol, and the like. A preferred group of carbohydrates includes lactose, trehalose, raffinose maltodextrins, and mannitol. Suitable polypeptides include aspartame. Amino acids include alanine and glycine, with glycine being preferred.

Additives, which are minor components of the composition of this invention, may be included for conformational stability during spray drying and for improving dispersibility of the powder. These additives include hydrophobic amino acids such as tryptophan, tyrosine, leucine, phenylalanine, and the like.

Suitable pH adjusters or buffers include organic salts prepared from organic acids and bases, such as sodium citrate, sodium ascorbate, and the like; sodium citrate is preferred.

Pulmonary administration of a micellar iRNA formulation may be achieved through metered dose spray devices with propellants such as tetrafluoroethane, heptafluoroethane, dimethylfluoropropane, tetrafluoropropane, butane, isobutane, dimethyl ether and other non-CFC and CFC propellants.

Oral or Nasal Delivery

For ease of exposition the formulations, compositions and methods in this section are discussed largely with regard to unmodified iRNA agents. It should be understood, however,

Attorney's Docket No.: 14174-072W01

that these formulations, compositions and methods can be practiced with other iRNA agents, *e.g.*, modified iRNA agents, and such practice is within the invention. Both the oral and nasal membranes offer advantages over other routes of administration. For example, drugs administered through these membranes have a rapid onset of action, provide therapeutic plasma levels, avoid first pass effect of hepatic metabolism, and avoid exposure of the drug to the hostile gastrointestinal (GI) environment. Additional advantages include easy access to the membrane sites so that the drug can be applied, localized and removed easily.

In oral delivery, compositions can be targeted to a surface of the oral cavity, *e.g.*, to sublingual mucosa which includes the membrane of ventral surface of the tongue and the floor of the mouth or the buccal mucosa which constitutes the lining of the cheek. The sublingual mucosa is relatively permeable thus giving rapid absorption and acceptable bioavailability of many drugs. Further, the sublingual mucosa is convenient, acceptable and easily accessible.

The ability of molecules to permeate through the oral mucosa appears to be related to molecular size, lipid solubility and peptide protein ionization. Small molecules, less than 1000 daltons appear to cross mucosa rapidly. As molecular size increases, the permeability decreases rapidly. Lipid soluble compounds are more permeable than non-lipid soluble molecules. Maximum absorption occurs when molecules are un-ionized or neutral in electrical charges. Therefore charged molecules present the biggest challenges to absorption through the oral mucosae.

A pharmaceutical composition of iRNA may also be administered to the buccal cavity of a human being by spraying into the cavity, without inhalation, from a metered dose spray dispenser, a mixed micellar pharmaceutical formulation as described above and a propellant. In one embodiment, the dispenser is first shaken prior to spraying the pharmaceutical formulation and propellant into the buccal cavity.

25 Devices

For ease of exposition the devices, formulations, compositions and methods in this section are discussed largely with regard to unmodified iRNA agents. It should be understood, however, that these devices, formulations, compositions and methods can be practiced with other

iRNA agents, *e.g.*, modified iRNA agents, and such practice is within the invention. An iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, (*e.g.*, a precursor, *e.g.*, a larger iRNA agent which can be processed into a sRNA agent, or a DNA which encodes an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, or precursor thereof) can be disposed on or in a device, *e.g.*, a device which implanted or otherwise placed in a subject. Exemplary devices include devices which are introduced into the vasculature, *e.g.*, devices inserted into the lumen of a vascular tissue, or which devices themselves form a part of the vasculature, including stents, catheters, heart valves, and other vascular devices. These devices, *e.g.*, catheters or stents, can be placed in the vasculature of the lung, heart, or leg.

Other devices include non-vascular devices, *e.g.*, devices implanted in the peritoneum, or in organ or glandular tissue, *e.g.*, artificial organs. The device can release a therapeutic substance in addition to a iRNA, *e.g.*, a device can release insulin.

Other devices include artificial joints, *e.g.*, hip joints, and other orthopedic implants.

In one embodiment, unit doses or measured doses of a composition that includes iRNA are dispensed by an implanted device. The device can include a sensor that monitors a parameter within a subject. For example, the device can include pump, *e.g.*, and, optionally, associated electronics.

Tissue, *e.g.*, cells or organs, such as the liver, can be treated with an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, (*e.g.*, a precursor, *e.g.*, a larger iRNA agent which can be processed into a sRNA agent, or a DNA which encodes an iRNA agent; *e.g.*, a double-stranded iRNA agent, or sRNA agent, or precursor thereof) *ex vivo* and then administered or implanted in a subject.

The tissue can be autologous, allogeneic, or xenogeneic tissue. For example, tissue (*e.g.*, liver) can be treated to reduce graft v. host disease. In other embodiments, the tissue is allogeneic and the tissue is treated to treat a disorder characterized by unwanted gene expression in that tissue, such as in the liver. In another example, tissue containing hematopoietic cells, *e.g.*, bone marrow hematopoietic cells, can be treated to inhibit unwanted cell proliferation.

Attorney's Docket No.: 14174-072W01

Introduction of treated tissue, whether autologous or transplant, can be combined with other therapies.

In some implementations, the iRNA treated cells are insulated from other cells, *e.g.*, by a semi-permeable porous barrier that prevents the cells from leaving the implant, but enables
5 molecules from the body to reach the cells and molecules produced by the cells to enter the body. In one embodiment, the porous barrier is formed from alginate.

In one embodiment, a contraceptive device is coated with or contains an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, (*e.g.*, a precursor, *e.g.*, a larger iRNA agent which can be processed into a sRNA agent, or a DNA which encodes an iRNA agent, *e.g.*, a
10 double-stranded iRNA agent, or sRNA agent, or precursor thereof). Exemplary devices include condoms, diaphragms, IUD (implantable uterine devices, sponges, vaginal sheaths, and birth control devices. In one embodiment, the iRNA is chosen to inactivate sperm or egg. In another embodiment, the iRNA is chosen to be complementary to a viral or pathogen RNA, *e.g.*, an RNA of an STD. In some instances, the iRNA composition can include a spermicide.

15 DOSAGE

In one aspect, the invention features a method of administering an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, to a subject (*e.g.*, a human subject). The method includes administering a unit dose of the iRNA agent, *e.g.*, a sRNA agent, *e.g.*, double stranded sRNA agent that (a) the double-stranded part is 19-25 nucleotides (nt) long, preferably 21-23 nt,
20 (b) is complementary to a target RNA (*e.g.*, an endogenous or pathogen target RNA), and, optionally, (c) includes at least one 3' overhang 1-5 nucleotide long. In one embodiment, the unit dose is less than 1.4 mg per kg of bodyweight, or less than 10, 5, 2, 1, 0.5, 0.1, 0.05, 0.01, 0.005, 0.001, 0.0005, 0.0001, 0.00005 or 0.00001 mg per kg of bodyweight, and less than 200 nmole of RNA agent (*e.g.* about 4.4×10^{16} copies) per kg of bodyweight, or less than 1500, 750, 300, 150,
25 75, 15, 7.5, 1.5, 0.75, 0.15, 0.075, 0.015, 0.0075, 0.0015, 0.00075, 0.00015 nmole of RNA agent per kg of bodyweight.

The defined amount can be an amount effective to treat or prevent a disease or disorder, *e.g.*, a disease or disorder associated with the target RNA, such as an RNA present in the liver.

The unit dose, for example, can be administered by injection (*e.g.*, intravenous or intramuscular), an inhaled dose, or a topical application. Particularly preferred dosages are less than 2, 1, or 0.1 mg/kg of body weight.

In a preferred embodiment, the unit dose is administered less frequently than once a day, *e.g.*, less than every 2, 4, 8 or 30 days. In another embodiment, the unit dose is not administered with a frequency (*e.g.*, not a regular frequency). For example, the unit dose may be administered a single time.

In one embodiment, the effective dose is administered with other traditional therapeutic modalities. In one embodiment, the subject has a viral infection and the modality is an antiviral agent other than an iRNA agent, *e.g.*, other than a double-stranded iRNA agent, or sRNA agent. In another embodiment, the subject has atherosclerosis and the effective dose of an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, is administered in combination with, *e.g.*, after surgical intervention, *e.g.*, angioplasty.

In one embodiment, a subject is administered an initial dose and one or more maintenance doses of an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, (*e.g.*, a precursor, *e.g.*, a larger iRNA agent which can be processed into a sRNA agent, or a DNA which encodes an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, or precursor thereof). The maintenance dose or doses are generally lower than the initial dose, *e.g.*, one-half less of the initial dose. A maintenance regimen can include treating the subject with a dose or doses ranging from 0.01 μ g to 1.4 mg/kg of body weight per day, *e.g.*, 10, 1, 0.1, 0.01, 0.001, or 0.00001 mg per kg of bodyweight per day. The maintenance doses are preferably administered no more than once every 5, 10, or 30 days. Further, the treatment regimen may last for a period of time which will vary depending upon the nature of the particular disease, its severity and the overall condition of the patient. In preferred embodiments the dosage may be delivered no more than once per day, *e.g.*, no more than once per 24, 36, 48, or more hours, *e.g.*, no more than once for every 5 or 8 days. Following treatment, the patient can be monitored for changes in his condition and for alleviation of the symptoms of the disease state. The dosage of the compound may either be increased in the event the patient does not respond significantly to current dosage

Attorney's Docket No.: 14174-072W01

levels, or the dose may be decreased if an alleviation of the symptoms of the disease state is observed, if the disease state has been ablated, or if undesired side-effects are observed.

The effective dose can be administered in a single dose or in two or more doses, as desired or considered appropriate under the specific circumstances. If desired to facilitate
5 repeated or frequent infusions, implantation of a delivery device, *e.g.*, a pump, semi-permanent stent (*e.g.*, intravenous, intraperitoneal, intracisternal or intracapsular), or reservoir may be advisable.

In one embodiment, the iRNA agent pharmaceutical composition includes a plurality of iRNA agent species. In another embodiment, the iRNA agent species has sequences that are
10 non-overlapping and non-adjacent to another species with respect to a naturally occurring target sequence. In another embodiment, the plurality of iRNA agent species is specific for different naturally occurring target genes. In another embodiment, the iRNA agent is allele specific.

In some cases, a patient is treated with a iRNA agent in conjunction with other therapeutic modalities. For example, a patient being treated for a liver disease can be
15 administered an iRNA agent specific for a target gene known to enhance the progression of the disease in conjunction with a drug known to inhibit activity of the target gene product. For example, a patient being treated for a cancer of the liver can be administered an iRNA agent specific for a target essential for tumor cell proliferation in conjunction with a chemotherapy.

Following successful treatment, it may be desirable to have the patient undergo
20 maintenance therapy to prevent the recurrence of the disease state, wherein the compound of the invention is administered in maintenance doses, ranging from 0.01 μg to 100 g per kg of body weight (see US 6,107,094).

The concentration of the iRNA agent composition is an amount sufficient to be effective in treating or preventing a disorder or to regulate a physiological condition in humans. The
25 concentration or amount of iRNA agent administered will depend on the parameters determined for the agent and the method of administration, *e.g.* nasal, buccal, pulmonary. For example, nasal formulations tend to require much lower concentrations of some ingredients in order to

avoid irritation or burning of the nasal passages. It is sometimes desirable to dilute an oral formulation up to 10-100 times in order to provide a suitable nasal formulation.

Certain factors may influence the dosage required to effectively treat a subject, including but not limited to the severity of the disease or disorder, previous treatments, the general health and/or age of the subject, and other diseases present. Moreover, treatment of a subject with a therapeutically effective amount of an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, (*e.g.*, a precursor, *e.g.*, a larger iRNA agent which can be processed into a sRNA agent, or a DNA which encodes an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, or precursor thereof) can include a single treatment or, preferably, can include a series of treatments. It will also be appreciated that the effective dosage of a iRNA agent such as a sRNA agent used for treatment may increase or decrease over the course of a particular treatment. Changes in dosage may result and become apparent from the results of diagnostic assays as described herein. For example, the subject can be monitored after administering a iRNA agent composition. Based on information from the monitoring, an additional amount of the iRNA agent composition can be administered.

Dosing is dependent on severity and responsiveness of the disease condition to be treated, with the course of treatment lasting from several days to several months, or until a cure is effected or a diminution of disease state is achieved. Optimal dosing schedules can be calculated from measurements of drug accumulation in the body of the patient. Persons of ordinary skill can easily determine optimum dosages, dosing methodologies and repetition rates. Optimum dosages may vary depending on the relative potency of individual compounds, and can generally be estimated based on EC50s found to be effective in *in vitro* and *in vivo* animal models. In some embodiments, the animal models include transgenic animals that express a human gene, *e.g.* a gene that produces a target RNA. The transgenic animal can be deficient for the corresponding endogenous RNA. In another embodiment, the composition for testing includes a iRNA agent that is complementary, at least in an internal region, to a sequence that is conserved between the target RNA in the animal model and the target RNA in a human.

The inventors have discovered that iRNA agents described herein can be administered to mammals, particularly large mammals such as nonhuman primates or humans in a number of ways.

In one embodiment, the administration of the iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, composition is parenteral, *e.g.* intravenous (*e.g.*, as a bolus or as a diffusible infusion), intradermal, intraperitoneal, intramuscular, intrathecal, intraventricular, intracranial, subcutaneous, transmucosal, buccal, sublingual, endoscopic, rectal, oral, vaginal, topical, pulmonary, intranasal, urethral or ocular. Administration can be provided by the subject or by another person, *e.g.*, a health care provider. The medication can be provided in measured doses or in a dispenser which delivers a metered dose. Selected modes of delivery are discussed in more detail below.

The invention provides methods, compositions, and kits, for rectal administration or delivery of iRNA agents described herein.

Accordingly, an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, (*e.g.*, a precursor, *e.g.*, a larger iRNA agent which can be processed into a sRNA agent, or a DNA which encodes a an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, or precursor thereof) described herein, *e.g.*, a therapeutically effective amount of a iRNA agent described herein, *e.g.*, a iRNA agent having a double stranded region of less than 40, and preferably less than 30 nucleotides and having one or two 1-3 nucleotide single strand 3' overhangs can be administered rectally, *e.g.*, introduced through the rectum into the lower or upper colon. This approach is particularly useful in the treatment of, inflammatory disorders, disorders characterized by unwanted cell proliferation, *e.g.*, polyps, or colon cancer.

The medication can be delivered to a site in the colon by introducing a dispensing device, *e.g.*, a flexible, camera-guided device similar to that used for inspection of the colon or removal of polyps, which includes means for delivery of the medication.

The rectal administration of the iRNA agent is by means of an enema. The iRNA agent of the enema can be dissolved in a saline or buffered solution. The rectal administration can also

by means of a suppository, which can include other ingredients, *e.g.*, an excipient, *e.g.*, cocoa butter or hydropropylmethylcellulose.

Any of the iRNA agents described herein can be administered orally, *e.g.*, in the form of tablets, capsules, gel capsules, lozenges, troches or liquid syrups. Further, the composition can be applied topically to a surface of the oral cavity.

Any of the iRNA agents described herein can be administered buccally. For example, the medication can be sprayed into the buccal cavity or applied directly, *e.g.*, in a liquid, solid, or gel form to a surface in the buccal cavity. This administration is particularly desirable for the treatment of inflammations of the buccal cavity, *e.g.*, the gums or tongue, *e.g.*, in one embodiment, the buccal administration is by spraying into the cavity, *e.g.*, without inhalation, from a dispenser, *e.g.*, a metered dose spray dispenser that dispenses the pharmaceutical composition and a propellant.

Any of the iRNA agents described herein can be administered to ocular tissue. For example, the medications can be applied to the surface of the eye or nearby tissue, *e.g.*, the inside of the eyelid. They can be applied topically, *e.g.*, by spraying, in drops, as an eyewash, or an ointment. Administration can be provided by the subject or by another person, *e.g.*, a health care provider. The medication can be provided in measured doses or in a dispenser which delivers a metered dose. The medication can also be administered to the interior of the eye, and can be introduced by a needle or other delivery device which can introduce it to a selected area or structure. Ocular treatment is particularly desirable for treating inflammation of the eye or nearby tissue.

Any of the iRNA agents described herein can be administered directly to the skin. For example, the medication can be applied topically or delivered in a layer of the skin, *e.g.*, by the use of a microneedle or a battery of microneedles which penetrate into the skin, but preferably not into the underlying muscle tissue. Administration of the iRNA agent composition can be topical. Topical applications can, for example, deliver the composition to the dermis or epidermis of a subject. Topical administration can be in the form of transdermal patches, ointments, lotions, creams, gels, drops, suppositories, sprays, liquids or powders. A composition for topical administration can be formulated as a liposome, micelle, emulsion, or other lipophilic

molecular assembly. The transdermal administration can be applied with at least one penetration enhancer, such as iontophoresis, phonophoresis, and sonophoresis.

Any of the iRNA agents described herein can be administered to the pulmonary system.

Pulmonary administration can be achieved by inhalation or by the introduction of a delivery device into the pulmonary system, *e.g.*, by introducing a delivery device which can dispense the medication. A preferred method of pulmonary delivery is by inhalation. The medication can be provided in a dispenser which delivers the medication, *e.g.*, wet or dry, in a form sufficiently small such that it can be inhaled. The device can deliver a metered dose of medication. The subject, or another person, can administer the medication.

Pulmonary delivery is effective not only for disorders which directly affect pulmonary tissue, but also for disorders which affect other tissue.

iRNA agents can be formulated as a liquid or nonliquid, *e.g.*, a powder, crystal, or aerosol for pulmonary delivery.

Any of the iRNA agents described herein can be administered nasally. Nasal administration can be achieved by introduction of a delivery device into the nose, *e.g.*, by introducing a delivery device which can dispense the medication. Methods of nasal delivery include spray, aerosol, liquid, *e.g.*, by drops, or by topical administration to a surface of the nasal cavity. The medication can be provided in a dispenser with delivery of the medication, *e.g.*, wet or dry, in a form sufficiently small such that it can be inhaled. The device can deliver a metered dose of medication. The subject, or another person, can administer the medication.

Nasal delivery is effective not only for disorders which directly affect nasal tissue, but also for disorders which affect other tissue

iRNA agents can be formulated as a liquid or nonliquid, *e.g.*, a powder, crystal, or for nasal delivery.

An iRNA agent can be packaged in a viral natural capsid or in a chemically or enzymatically produced artificial capsid or structure derived therefrom.

The dosage of a pharmaceutical composition including a iRNA agent can be administered in order to alleviate the symptoms of a disease state, *e.g.*, cancer or a cardiovascular disease. A subject can be treated with the pharmaceutical composition by any of the methods mentioned above.

- 5 Gene expression in a subject can be modulated by administering a pharmaceutical composition including an iRNA agent.

A subject can be treated by administering a defined amount of an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, (*e.g.*, a precursor, *e.g.*, a larger iRNA agent which can be processed into a sRNA agent) composition that is in a powdered form, *e.g.*, a collection of
10 microparticles, such as crystalline particles. The composition can include a plurality of iRNA agents, *e.g.*, specific for one or more different endogenous target RNAs. The method can include other features described herein.

A subject can be treated by administering a defined amount of an iRNA agent composition that is prepared by a method that includes spray-drying, *i.e.* atomizing a liquid
15 solution, emulsion, or suspension, immediately exposing the droplets to a drying gas, and collecting the resulting porous powder particles. The composition can include a plurality of iRNA agents, *e.g.*, specific for one or more different endogenous target RNAs. The method can include other features described herein.

The iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, (*e.g.*, a precursor, *e.g.*, a larger iRNA agent which can be processed into a sRNA agent, or a DNA which encodes
20 an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, or precursor thereof), can be provided in a powdered, crystallized or other finely divided form, with or without a carrier, *e.g.*, a micro- or nano-particle suitable for inhalation or other pulmonary delivery. This can include providing an aerosol preparation, *e.g.*, an aerosolized spray-dried composition. The aerosol
25 composition can be provided in and/or dispensed by a metered dose delivery device.

The subject can be treated for a condition treatable by inhalation, *e.g.*, by aerosolizing a spray-dried iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, (*e.g.*, a precursor, *e.g.*, a larger iRNA agent which can be processed into a sRNA agent, or a DNA which encodes

an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, or precursor thereof) composition and inhaling the aerosolized composition. The iRNA agent can be an sRNA. The composition can include a plurality of iRNA agents, *e.g.*, specific for one or more different endogenous target RNAs. The method can include other features described herein.

- 5 A subject can be treated by, for example, administering a composition including an effective/defined amount of an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, (*e.g.*, a precursor, *e.g.*, a larger iRNA agent which can be processed into a sRNA agent, or a DNA which encodes an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, or precursor thereof), wherein the composition is prepared by a method that includes spray-drying, 10 lyophilization, vacuum drying, evaporation, fluid bed drying, or a combination of these techniques

- In another aspect, the invention features a method that includes: evaluating a parameter related to the abundance of a transcript in a cell of a subject; comparing the evaluated parameter to a reference value; and if the evaluated parameter has a preselected relationship to the reference 15 value (*e.g.*, it is greater), administering a iRNA agent (or a precursor, *e.g.*, a larger iRNA agent which can be processed into a sRNA agent, or a DNA which encodes a iRNA agent or precursor thereof) to the subject. In one embodiment, the iRNA agent includes a sequence that is complementary to the evaluated transcript. For example, the parameter can be a direct measure of transcript levels, a measure of a protein level, a disease or disorder symptom or 20 characterization (*e.g.*, rate of cell proliferation and/or tumor mass, viral load).

- In another aspect, the invention features a method that includes: administering a first amount of a composition that comprises an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, (*e.g.*, a precursor, *e.g.*, a larger iRNA agent which can be processed into a sRNA agent, or a DNA which encodes an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA 25 agent, or precursor thereof) to a subject, wherein the iRNA agent includes a strand substantially complementary to a target nucleic acid; evaluating an activity associated with a protein encoded by the target nucleic acid; wherein the evaluation is used to determine if a second amount should be administered. In a preferred embodiment the method includes administering a second amount

Attorney's Docket No.: 14174-072W01

of the composition, wherein the timing of administration or dosage of the second amount is a function of the evaluating. The method can include other features described herein.

In another aspect, the invention features a method of administering a source of a double-stranded iRNA agent (ds iRNA agent) to a subject. The method includes administering or
5 implanting a source of a ds iRNA agent, *e.g.*, a sRNA agent, that (a) includes a double-stranded region that is 19-25 nucleotides long, preferably 21-23 nucleotides, (b) is complementary to a target RNA (*e.g.*, an endogenous RNA or a pathogen RNA), and, optionally, (c) includes at least one 3' overhang 1-5 nt long. In one embodiment, the source releases ds iRNA agent over time,
10 *e.g.* the source is a controlled or a slow release source, *e.g.*, a microparticle that gradually releases the ds iRNA agent. In another embodiment, the source is a pump, *e.g.*, a pump that includes a sensor or a pump that can release one or more unit doses.

In one aspect, the invention features a pharmaceutical composition that includes an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, (*e.g.*, a precursor, *e.g.*, a larger iRNA agent which can be processed into a sRNA agent, or a DNA which encodes an iRNA agent, *e.g.*,
15 a double-stranded iRNA agent, or sRNA agent, or precursor thereof) including a nucleotide sequence complementary to a target RNA, *e.g.*, substantially and/or exactly complementary. The target RNA can be a transcript of an endogenous human gene. In one embodiment, the iRNA agent (a) is 19-25 nucleotides long, preferably 21-23 nucleotides, (b) is complementary to an endogenous target RNA, and, optionally, (c) includes at least one 3' overhang 1-5 nt long. In one
20 embodiment, the pharmaceutical composition can be an emulsion, microemulsion, cream, jelly, or liposome.

In one example the pharmaceutical composition includes an iRNA agent mixed with a topical delivery agent. The topical delivery agent can be a plurality of microscopic vesicles. The microscopic vesicles can be liposomes. In a preferred embodiment the liposomes are cationic
25 liposomes.

In another aspect, the pharmaceutical composition includes an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent (*e.g.*, a precursor, *e.g.*, a larger iRNA agent which can be processed into a sRNA agent, or a DNA which encodes an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, or precursor thereof) admixed with a topical penetration

Attorney's Docket No.: 14174-072W01

enhancer. In one embodiment, the topical penetration enhancer is a fatty acid. The fatty acid can be arachidonic acid, oleic acid, lauric acid, caprylic acid, capric acid, myristic acid, palmitic acid, stearic acid, linoleic acid, linolenic acid, dicaprate, tricaprate, monoolein, dilaurin, glyceryl 1-monocaprate, 1-dodecylazacycloheptan-2-one, an acylcarnitine, an acylcholine, or a C₁₋₁₀ alkyl ester, monoglyceride, diglyceride or pharmaceutically acceptable salt thereof.

In another embodiment, the topical penetration enhancer is a bile salt. The bile salt can be cholic acid, dehydrocholic acid, deoxycholic acid, gluchoic acid, glycholic acid, glycodeoxycholic acid, taurocholic acid, taurodeoxycholic acid, chenodeoxycholic acid, ursodeoxycholic acid, sodium tauro-24,25-dihydro-fusidate, sodium glycodihydrofusidate, polyoxyethylene-9-lauryl ether or a pharmaceutically acceptable salt thereof.

In another embodiment, the penetration enhancer is a chelating agent. The chelating agent can be EDTA, citric acid, a salicylate, a N-acyl derivative of collagen, laureth-9, an N-amino acyl derivative of a beta-diketone or a mixture thereof.

In another embodiment, the penetration enhancer is a surfactant, *e.g.*, an ionic or nonionic surfactant. The surfactant can be sodium lauryl sulfate, polyoxyethylene-9-lauryl ether, polyoxyethylene-20-cetyl ether, a perfluorchemical emulsion or mixture thereof.

In another embodiment, the penetration enhancer can be selected from a group consisting of unsaturated cyclic ureas, 1-alkyl-alkones, 1-alkenylazacyclo-alkanones, steroidal anti-inflammatory agents and mixtures thereof. In yet another embodiment the penetration enhancer can be a glycol, a pyrrol, an azone, or a terpenes.

In one aspect, the invention features a pharmaceutical composition including an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, (*e.g.*, a precursor, *e.g.*, a larger iRNA agent which can be processed into a sRNA agent, or a DNA which encodes an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, or precursor thereof) in a form suitable for oral delivery. In one embodiment, oral delivery can be used to deliver an iRNA agent composition to a cell or a region of the gastro-intestinal tract, *e.g.*, small intestine, colon (*e.g.*, to treat a colon cancer), and so forth. The oral delivery form can be tablets, capsules or gel capsules. In one embodiment, the iRNA agent of the pharmaceutical composition modulates expression of a

Attorney's Docket No.: 14174-072W01

cellular adhesion protein, modulates a rate of cellular proliferation, or has biological activity against eukaryotic pathogens or retroviruses. In another embodiment, the pharmaceutical composition includes an enteric material that substantially prevents dissolution of the tablets, capsules or gel capsules in a mammalian stomach. In a preferred embodiment the enteric material is a coating. The coating can be acetate phthalate, propylene glycol, sorbitan monoleate, 5 cellulose acetate trimellitate, hydroxy propyl methylcellulose phthalate or cellulose acetate phthalate.

In another embodiment, the oral dosage form of the pharmaceutical composition includes a penetration enhancer. The penetration enhancer can be a bile salt or a fatty acid. The bile salt 10 can be ursodeoxycholic acid, chenodeoxycholic acid, and salts thereof. The fatty acid can be capric acid, lauric acid, and salts thereof.

In another embodiment, the oral dosage form of the pharmaceutical composition includes an excipient. In one example the excipient is polyethyleneglycol. In another example the excipient is precirol.

15 In another embodiment, the oral dosage form of the pharmaceutical composition includes a plasticizer. The plasticizer can be diethyl phthalate, triacetin dibutyl sebacate, dibutyl phthalate or triethyl citrate.

In one aspect, the invention features a pharmaceutical composition including an iRNA agent and a delivery vehicle. In one embodiment, the iRNA agent is (a) is 19-25 nucleotides 20 long, preferably 21-23 nucleotides, (b) is complementary to an endogenous target RNA, and, optionally, (c) includes at least one 3' overhang 1-5 nucleotides long.

In one embodiment, the delivery vehicle can deliver an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, (*e.g.*, a precursor, *e.g.*, a larger iRNA agent which can be processed into a sRNA agent, or a DNA which encodes an iRNA agent, *e.g.*, a double-stranded 25 iRNA agent, or sRNA agent, or precursor thereof) to a cell by a topical route of administration. The delivery vehicle can be microscopic vesicles. In one example the microscopic vesicles are liposomes. In a preferred embodiment the liposomes are cationic liposomes. In another example the microscopic vesicles are micelles. In one aspect, the invention features a pharmaceutical

Attorney's Docket No.: 14174-072W01

composition including an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, (*e.g.*, a precursor, *e.g.*, a larger iRNA agent which can be processed into a sRNA agent, or a DNA which encodes an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, or precursor thereof) in an injectable dosage form. In one embodiment, the injectable dosage form of the
5 pharmaceutical composition includes sterile aqueous solutions or dispersions and sterile powders. In a preferred embodiment the sterile solution can include a diluent such as water, saline solution; fixed oils, polyethylene glycols, glycerin, or propylene glycol.

In one aspect, the invention features a pharmaceutical composition including an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, (*e.g.*, a precursor, *e.g.*, a larger iRNA
10 agent which can be processed into a sRNA agent, or a DNA which encodes an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, or precursor thereof) in oral dosage form. In one embodiment, the oral dosage form is selected from the group consisting of tablets, capsules and gel capsules. In another embodiment, the pharmaceutical composition includes an enteric material that substantially prevents dissolution of the tablets, capsules or gel capsules in a
15 mammalian stomach. In a preferred embodiment the enteric material is a coating. The coating can be acetate phthalate, propylene glycol, sorbitan monooleate, cellulose acetate trimellitate, hydroxy propyl methyl cellulose phthalate or cellulose acetate phthalate. In one embodiment, the oral dosage form of the pharmaceutical composition includes a penetration enhancer, *e.g.*, a penetration enhancer described herein.

In one aspect, the invention features a pharmaceutical composition including an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, (*e.g.*, a precursor, *e.g.*, a larger iRNA agent which can be processed into a sRNA agent, or a DNA which encodes an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, or precursor thereof) in a rectal dosage form. In
20 one embodiment, the rectal dosage form is an enema. In another embodiment, the rectal dosage form is a suppository.
25

In one aspect, the invention features a pharmaceutical composition including an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, (*e.g.*, a precursor, *e.g.*, a larger iRNA agent which can be processed into a sRNA agent, or a DNA which encodes an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, or precursor thereof) in a vaginal dosage form.

Attorney's Docket No.: 14174-072W01

In one embodiment, the vaginal dosage form is a suppository. In another embodiment, the vaginal dosage form is a foam, cream, or gel.

In one aspect, the invention features a pharmaceutical composition including an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, (*e.g.*, a precursor, *e.g.*, a larger iRNA agent which can be processed into a sRNA agent, or a DNA which encodes an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, or precursor thereof) in a pulmonary or nasal dosage form. In one embodiment, the iRNA agent is incorporated into a particle, *e.g.*, a macroparticle, *e.g.*, a microsphere. The particle can be produced by spray drying, lyophilization, evaporation, fluid bed drying, vacuum drying, or a combination thereof. The microsphere can be formulated as a suspension, a powder, or an implantable solid.

In one aspect, the invention features a spray-dried iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, (*e.g.*, a precursor, *e.g.*, a larger iRNA agent which can be processed into a sRNA agent, or a DNA which encodes an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, or precursor thereof) composition suitable for inhalation by a subject, including: (a) a therapeutically effective amount of a iRNA agent suitable for treating a condition in the subject by inhalation; (b) a pharmaceutically acceptable excipient selected from the group consisting of carbohydrates and amino acids; and (c) optionally, a dispersibility-enhancing amount of a physiologically-acceptable, water-soluble polypeptide.

In one embodiment, the excipient is a carbohydrate. The carbohydrate can be selected from the group consisting of monosaccharides, disaccharides, trisaccharides, and polysaccharides. In a preferred embodiment the carbohydrate is a monosaccharide selected from the group consisting of dextrose, galactose, mannitol, D-mannose, sorbitol, and sorbose. In another preferred embodiment the carbohydrate is a disaccharide selected from the group consisting of lactose, maltose, sucrose, and trehalose.

In another embodiment, the excipient is an amino acid. In one embodiment, the amino acid is a hydrophobic amino acid. In a preferred embodiment the hydrophobic amino acid is selected from the group consisting of alanine, isoleucine, leucine, methionine, phenylalanine, proline, tryptophan, and valine. In yet another embodiment the amino acid is a polar amino acid. In a preferred embodiment the amino acid is selected from the group consisting of arginine,

Attorney's Docket No.: 14174-072W01

histidine, lysine, cysteine, glycine, glutamine, serine, threonine, tyrosine, aspartic acid and glutamic acid.

In one embodiment, the dispersibility-enhancing polypeptide is selected from the group consisting of human serum albumin, α -lactalbumin, trypsinogen, and polyalanine.

5 In one embodiment, the spray-dried iRNA agent composition includes particles having a mass median diameter (MMD) of less than 10 microns. In another embodiment, the spray-dried iRNA agent composition includes particles having a mass median diameter of less than 5 microns. In yet another embodiment the spray-dried iRNA agent composition includes particles having a mass median aerodynamic diameter (MMAD) of less than 5 microns.

10 In certain other aspects, the invention provides kits that include a suitable container containing a pharmaceutical formulation of an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, (*e.g.*, a precursor, *e.g.*, a larger iRNA agent which can be processed into a sRNA agent, or a DNA which encodes an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, or precursor thereof). In certain embodiments the individual components of the
15 pharmaceutical formulation may be provided in one container. Alternatively, it may be desirable to provide the components of the pharmaceutical formulation separately in two or more containers, *e.g.*, one container for an iRNA agent preparation, and at least another for a carrier compound. The kit may be packaged in a number of different configurations such as one or more containers in a single box. The different components can be combined, *e.g.*, according to
20 instructions provided with the kit. The components can be combined according to a method described herein, *e.g.*, to prepare and administer a pharmaceutical composition. The kit can also include a delivery device.

In another aspect, the invention features a device, *e.g.*, an implantable device, wherein the device can dispense or administer a composition that includes an iRNA agent, *e.g.*, a double-
25 stranded iRNA agent, or sRNA agent, (*e.g.*, a precursor, *e.g.*, a larger iRNA agent which can be processed into a sRNA agent, or a DNA which encodes an iRNA agent, *e.g.*, a double-stranded iRNA agent, or sRNA agent, or precursor thereof), *e.g.*, a iRNA agent that silences an endogenous transcript. In one embodiment, the device is coated with the composition. In another embodiment the iRNA agent is disposed within the device. In another embodiment, the

Attorney's Docket No.: 14174-072W01

device includes a mechanism to dispense a unit dose of the composition. In other embodiments the device releases the composition continuously, *e.g.*, by diffusion. Exemplary devices include stents, catheters, pumps, artificial organs or organ components (*e.g.*, artificial heart, a heart valve, etc.), and sutures.

5 As used herein, the term "crystalline" describes a solid having the structure or characteristics of a crystal, *i.e.*, particles of three-dimensional structure in which the plane faces intersect at definite angles and in which there is a regular internal structure. The compositions of the invention may have different crystalline forms. Crystalline forms can be prepared by a variety of methods, including, for example, spray drying.

10 The invention is further illustrated by the following examples, which should not be construed as further limiting.

EXAMPLES

Example 1: apoB protein as a therapeutic target for lipid-based diseases

15 Apolipoprotein B (apoB) is a candidate target gene for the development of novel therapies for lipid-based diseases.

Methods described herein can be used to evaluate the efficacy of a particular siRNA as a therapeutic tool for treating lipid metabolism disorders resulting elevated apoB levels. Use of siRNA duplexes to selectively bind and inactivate the target apoB mRNA is an approach to treat these disorders.

20 Two approaches:

- i) Inhibition of apoB in *ex-vivo* models by transfecting siRNA duplexes homologous to human apoB mRNA in a human hepatoma cell line (Hep G2) and monitor the level of the protein and the RNA using the Western blotting and RT-PCR methods, respectively. siRNA molecules that efficiently inhibit apoB expression will be tested for similar effects *in vivo*.
- 25 ii) *In vivo* trials using an apoB transgenic mouse model (apoB100 Transgenic Mice, C57BL/6NTac-TgN (APOB100), Order Model #s:1004-T (hemizygotes), B6 (control)). siRNA duplexes are designed to target apoB-100 or CETP/apoB double transgenic mice which express both cholesteryl ester transfer protein (CETP) and apoB. The effect of the siRNA on gene expression *in vivo* can be measured by monitoring the HDL/LDL cholesterol level in serum. The

Attorney's Docket No.: 14174-072W01

results of these experiments would indicate the therapeutic potential of siRNAs to treat lipid-based diseases, including hypercholesterolemia, HDL/LDL cholesterol imbalance, familial combined hyperlipidemia, and acquired hyperlipidemia.

- 5 *Background* Fats, in the form of triglycerides, are ideal for energy storage because they are highly reduced and anhydrous. An adipocyte (or fat cell) consists of a nucleus, a cell membrane, and triglycerides, and its function is to store triglycerides.

- The lipid portion of the human diet consists largely of triglycerides and cholesterol (and its esters). These must be emulsified and digested to be absorbed. Specifically, fats
10 (triacylglycerols) are ingested. Bile (bile acids, salts, and cholesterol), which is made in the liver, is secreted by the gall bladder. Pancreatic lipase digests the triglycerides to fatty acids, and also digests di-, and mono-acylglycerols, which are absorbed by intestinal epithelial cells and then are resynthesized into triacylglycerols once inside the cells. These triglycerides and some
15 cholesterol are combined with apolipoproteins to produce chylomicrons. Chylomicrons consist of approximately 95% triglycerides. The chylomicrons transport fatty acids to peripheral tissues. Any excess fat is stored in adipose tissue.

- Lipid transport and clearance from the blood into cells, and from the cells into the blood and the liver, is mediated by the lipoprotein transport proteins. This class of approximately 17 proteins can be divided into three groups: Apolipoproteins, lipoprotein processing proteins, and
20 lipoprotein receptors.

- Apolipoproteins coat lipoprotein particles, and include the A-I, A-II, A-IV, B, CI, CII, CIII, D, E, Apo(a) proteins. Lipoprotein processing proteins include lipoprotein lipase, hepatic lipase, lecithin cholesterol acyltransferase and cholesterol ester transfer protein. Lipoprotein receptors include the low density lipoprotein (LDL) receptor, chylomicron-remnant receptor (the
25 LDL receptor like protein or LDL receptor related protein - LRP) and the scavenger receptor.

- Lipoprotein Metabolism* Since the triglycerides, cholesterol esters, and cholesterol absorbed into the small intestine are not soluble in aqueous medium, they must be combined with suitable proteins (apolipoproteins) in order to prevent them from forming large oil droplets. The resulting
30 lipoproteins undergo a type of metabolism as they pass through the bloodstream and certain organs (notably the liver).

Attorney's Docket No.: 14174-072W01

Also synthesized in the liver is high density lipoprotein (HDL), which contains the apoproteins A-1, A-2, C-1, and D; HDL collects cholesterol from peripheral tissues and blood vessels and returns it to the liver. LDL is taken up by specific cell surface receptors into an endosome, which fuses with a lysosome where cholesterol ester is converted to free cholesterol.

- 5 The apoproteins (including apo B-100) are digested to amino acids. The receptor protein is recycled to the cell membrane.

The free cholesterol formed by this process has two fates. First, it can move to the endoplasmic reticulum (ER), where it can inhibit HMG-CoA reductase, the synthesis of HMG-CoA reductase, and the synthesis of cell surface receptors for LDL. Also in the ER, cholesterol
10 can speed up the degradation of HMG-CoA reductase. The free cholesterol can also be converted by acyl-CoA and acyl transferase (ACAT) to cholesterol esters, which form oil droplets.

ApoB is the major apolipoprotein of chylomicrons of very low density lipoproteins (VLDL, which carry most of the plasma triglyceride) and low density lipoprotein (LDL, which
15 carry most of the plasma cholesterol). ApoB exists in human plasma in two isoforms, apoB-48 and apoB-100.

ApoB-100 is the major physiological ligand for the LDL receptor. The ApoB precursor has 4563 amino acids, and the mature apoB-100 has 4536 amino acid residues. The LDL-binding domain of ApoB-100 is proposed to be located between residues 3129 and 3532. ApoB-100 is
20 synthesized in the liver and is required for the assembly of very low density lipoproteins VLDL and for the preparation of apoB-100 to transport triglycerides (TG) and cholesterol from the liver to other tissues. ApoB-100 does not interchange between lipoprotein particles, as do the other lipoproteins, and it is found in IDL and LDL particles. After the removal of apolipoproteins A, E and C, apoB is incorporation into VLDL by hepatocytes. ApoB-48 is present in chylomicrons
25 and plays an essential role in the intestinal absorption of dietary fats. ApoB-48 is synthesized in the small intestine. It comprises the N-terminal 48% of apoB-100 and is produced by a posttranscriptional apoB-100 mRNA editing event at codon 2153 (C to U). This editing event is a product of the apoBEC-1b enzyme, which is expressed in the intestine. This editing event creates a stop codon instead of a glutamine codon, and therefore apoB-48, instead of apoB-100 is
30 expressed in the intestine (apoB-100 is expressed in the liver).

Attorney's Docket No.: 14174-072W01

There is also strong evidence that plasma apoB levels may be a better index of the risk of coronary artery disease (CAD) than total or LDL cholesterol levels. Clinical studies have demonstrated the value of measuring apoB in hypertriglyceridemic, hypercholesterolemic and normalipidemic subjects.

Table 4. Reference Range Lipid level in the Blood

Lipid	Range (mmols/ L)
Plasma Cholesterol	3.5-6.5
Low density lipoprotein	1.55-4.4
Very low density lipoprotein	0.128-0.645
High density lipoprotein/ triglycerides	0.5-2.1
Total lipid	4.0-10g / L

Molecular genetics of lipid metabolism in both humans and induced mutant mouse models

- 5 Elevated plasma levels of LDL and apoB are associated with a higher risk for atherosclerosis and coronary heart disease, a leading cause of mortality. ApoB is the mandatory constituent of LDL particles. In addition to its role in lipoprotein metabolism, apoB has also been implicated as a factor in male infertility and fetal development. Furthermore, two quantitative trait loci regulating plasma apoB levels have been discovered, through the use of transgenic mouse
- 10 models. Future experiments will facilitate the identification of human orthologous genes encoding regulators of plasma apoB levels. These loci are candidate therapeutic targets for human disorders characterized by altered plasma apoB levels. Such disorders include non-apoB linked hypobetalipoproteinemia and familial combined hyperlipidemia. The identification of these genetic loci would also reveal possible new pathways involved in the regulation of apoB
- 15 secretion, potentially providing novel sites for pharmacological therapy.

Diseases and Clinical Pharmacology Familial combined hyperlipemia (FCHL) affects an estimated one in 10 Americans. FCHL can cause premature heart disease.

- Familial Hypercholesterolemia (high level of apo B)* A common genetic disorder of lipid
- 20 metabolism. Familial hypercholesterolemia is characterized by elevated serum TC in association with xanthelasma, tendon and tuberous xanthomas, accelerated atherosclerosis, and early death from myocardial infarction (MI). It is caused by absent or defective LDL cell receptors, resulting in delayed LDL clearance, an increase in plasma LDL levels, and an accumulation of LDL cholesterol in macrophages over joints and pressure points, and in blood vessels.

Attorney's Docket No.: 14174-072W01

Atherosclerosis (high level of apo B) Atherosclerosis develops as a deposition of cholesterol and fat in the arterial wall due to disturbances in lipid transport and clearance from the blood into cells and from the cells to blood and the liver.

Clinical studies have demonstrated that elevation of total cholesterol (TC), low-density lipoprotein cholesterol (LDL-C) and apoB-100 promote human atherosclerosis. Similarly, decreased levels of high-density lipoprotein cholesterol (HDL-C) are associated with the development of atherosclerosis.

ApoB may be a factor in the genetic cause of high cholesterol.

The risk of coronary artery disease (CAD) (high level of apo B) Cardiovascular disease, including coronary heart disease and stroke, is a leading cause of death and disability. The major risk factors include age, gender, elevated low-density lipoprotein cholesterol blood levels, decreased high-density lipoprotein cholesterol levels, cigarette smoking, hypertension, and diabetes. Emerging risk factors include elevated lipoprotein (a), remnant lipoproteins, and C reactive protein. Dietary intake, physical activity and genetics also impact cardiovascular risk. Hypertension and age are the major risk factors for stroke.

Abetalipoproteinemia, an inherited human disease characterized by a near-complete absence of apoB-containing lipoproteins in the plasma, is caused by mutations in the gene for microsomal triglyceride transfer protein (MTP).

Model for human atherosclerosis (Lipoprotein A transgenic mouse) Numerous studies have demonstrated that an elevated plasma level of lipoprotein(a) (Lp(a)) is a major independent risk factor for coronary heart disease (CHD). Current therapies, however, have little or no effect on apo(a) levels and the homology between apo(a) and plasminogen presents barriers to drug development. Lp(a) particles consist of apo(a) and apoB-100 proteins, and they are found only in primates and the hedgehog. The development of LPA transgenic mouse requires the creation of animals that express both human apoB and apo(a) transgenes to achieve assembly of Lp(a). An atherosclerosis mouse model would facilitate the study of the disease process and factors influencing it, and further would facilitate the development of therapeutic or preventive agents. There are several strategies for gene-oriented therapy. For example, the missing or non-functional gene can be replaced, or unwanted gene activity can be inhibited.

Model for lipid Metabolism and Atherosclerosis DNX Transgenic Sciences has demonstrated that both CETP/ApoB and ApoB transgenic mice develop atherosclerotic plaques.

- 5 *Model for apoB-100 overexpression* The apoB-100 transgenic mice express high levels of human apoB-100. They consequently demonstrate elevated serum levels of LDL cholesterol. After 6 months on a high-fat diet, the mice develop significant foam cell accumulation under the endothelium and within the media, as well as cholesterol crystals and fibrotic lesions.
- 10 *Model for Cholesteryl ester transfer protein over expression* The apoB-100 transgenic mice express the human enzyme, CETP, and consequently demonstrate a dramatically reduced level of serum HDL cholesterol.
- Model for apoB-100 and CETP overexpression* The apoB-100 transgenic mice express both
- 15 CETP and apoB-100, resulting in mice with a human like serum HDL/LDL distribution. Following 6 months on a high-fat diet these mice develop significant foam cell accumulation underlying the endothelium and within the media, as well as cholesterol crystals and fibrotic lesions.
- 20 *ApoB100 Transgenic Mice (Order Model #'s: 1004-T (hemizygotes), B6 (control))* These mice express high levels of human apoB-100, resulting in mice with elevated serum levels of LDL cholesterol. These mice are useful in identifying and evaluating compounds to reduce elevated levels of LDL cholesterol and the risk of atherosclerosis. When fed a high fat cholesterol diet, these mice develop significant foam cell accumulation under the endothelium
- 25 and within the media, and have significantly more complex atherosclerotic lesions than control animals.

- Double Transgenic Mice, CETP/ApoB100 (Order Model #: 1007-TT)* These mice express both CETP and apoB-100, resulting in a human-like serum HDL/LDL distribution. These mice are
- 30 useful for evaluating compounds to treat hypercholesterolemia or HDL/LDL cholesterol imbalance to reduce the risk of developing atherosclerosis. When fed a high fat high cholesterol

Attorney's Docket No.: 14174-072W01

diet, these mice develop significant foam cell accumulation underlying the endothelium and within the media, and have significantly more complex atherosclerotic lesions than control animals.

- 5 *ApoE gene knockout mouse* Homozygous apoE knockout mice exhibit strong hypercholesterolemia, primarily due to elevated levels of VLDL and IDL caused by a defect in lipoprotein clearance from plasma. These mice develop atherosclerotic lesions which progress with age and resemble human lesions (Zhang *et al.*, *Science* 258:46-71, 1992; Plump *et al.*, *Cell* 71:343-353, 1992; Nakashima *et al.*, *Arterioscler Thromp.* 14:133-140, 1994; Reddick *et al.*,
10 *Arterioscler Tromb.* 14:141-147, 1994). These mice are a promising model for studying the effect of diet and drugs on atherosclerosis.

- Low density lipoprotein receptor (LDLR) mediates lipoprotein clearance from plasma through the recognition of apoB and apoE on the surface of lipoprotein particles. Humans, who lack or have a decreased number of the LDL receptors, have familial hypercholesterolemia and
15 develop CHD at an early age.

- ApoE Knockout Mice (Order Model #: APOE-M)* The apoE knockout mouse was created by gene targeting in embryonic stem cells to disrupt the apoE gene. ApoE, a glycoprotein, is a structural component of very low density lipoprotein (VLDL) synthesized by the liver and
20 intestinally synthesized chylomicrons. It is also a constituent of a subclass of high density lipoproteins (HDLs) involved in cholesterol transport activity among cells. One of the most important roles of apoE is to mediate high affinity binding of chylomicrons and VLDL particles that contain apoE to the low density lipoprotein (LDL) receptor. This allows for the specific uptake of these particles by the liver which is necessary for transport preventing the
25 accumulation in plasma of cholesterol-rich remnants. The homozygous inactivation of the apoE gene results in animals that are devoid of apoE in their sera. The mice appear to develop normally, but they exhibit five times the normal serum plasma cholesterol and spontaneous atherosclerotic lesions. This is similar to a disease in people who have a variant form of the apoE gene that is defective in binding to the LDL receptor and are at risk for early development
30 of atherosclerosis and increased plasma triglyceride and cholesterol levels. There are indications that apoE is also involved in immune system regulation, nerve regeneration and muscle

Attorney's Docket No.: 14174-072W01

differentiation. The apoE knockout mice can be used to study the role of apoE in lipid metabolism, atherogenesis, and nerve injury, and to investigate intervention therapies that modify the atherogenic process.

ApoE4 Targeted Replacement Mouse (Order Model #: 001549-M) ApoE is a plasma protein involved in cholesterol transport, and the three human isoforms (E2, E3, and E4) have been associated with atherosclerosis and Alzheimer's disease. Gene targeting of 129 ES cells was used to replace the coding sequence of mouse apoE with human APOE4 without disturbing the murine regulatory sequences. The E4 isoform occurs in approximately 14% of the human population and is associated with increased plasma cholesterol and a greater risk of coronary artery disease. The Taconic apoE4 Targeted Replacement model has normal plasma cholesterol and triglyceride levels, but altered quantities of different plasma lipoprotein particles. This model also has delayed plasma clearance of cholesterol-rich lipoprotein particles (VLDL), with only half the clearance rate seen in the apoE3 Targeted Replacement model. Like the apoE3 model, the apoE4 mice develop altered plasma lipoprotein values and atherosclerotic plaques on an atherogenic diet. However, the atherosclerosis is more severe in the apoE4 model, with larger plaques and cholesterol apoE and apoB-48 levels twice that seen in the apoE3 model. The Taconic apoE4 Targeted Replacement model, along with the apoE2 and apoE3 Targeted Replacement Mice, provide an excellent tool for *in vivo* study of the human apoE isoforms.

CETP Transgenic Mice (Order Model #: 1003-T) These animals express the human plasma enzyme, CETP, resulting in mice with a dramatic reduction in serum HDL cholesterol. The mice can be useful in identifying and evaluating compounds that increase the levels of HDL cholesterol for reducing the risk of developing atherosclerosis

Transgene/Promoter: human apolipoprotein A-I These mice produce mouse HDL cholesterol particles that contain human apolipoprotein A-I. Transgenic expression is life-long in both sexes (Biochemical Genetics and Metabolism Laboratory, Rockefeller University, NY City).

A Mouse Model for Abetalipoproteinemia Abetalipoproteinemia, an inherited human disease characterized by a near-complete absence of apoB-containing lipoproteins in the plasma, is caused by mutations in the gene for microsomal triglyceride transfer protein (MTP). Gene

Attorney's Docket No.: 14174-072W01

targeting was used to knock out the mouse MTP gene (*Mtp*). In heterozygous knockout mice (*Mtp*^{+/-}), the MTP mRNA, protein, and activity levels were reduced by 50% in both liver and intestine. Recent studies with heterozygous MTP knockout mice have suggested that half-normal levels of MTP in the liver reduce apoB secretion. They hypothesized that reduced apoB secretion in the setting of half-normal MTP levels might be caused by a reduced MTP:apoB ratio in the endoplasmic reticulum, which would reduce the number of apoB-MTP interactions. If this hypothesis were true, half-normal levels of MTP might have little impact on lipoprotein secretion in the setting of half-normal levels of apoB synthesis (since the ratio of MTP to apoB would not be abnormally low) and might cause an exaggerated reduction in lipoprotein secretion in the setting of apoB overexpression (since the ratio of MTP to apoB would be even lower). To test this hypothesis, they examined the effects of heterozygous MTP deficiency on apoB metabolism in the setting of normal levels of apoB synthesis, half-normal levels of apoB synthesis (heterozygous *ApoB* deficiency), and increased levels of apoB synthesis (transgenic overexpression of human apoB). Contrary to their expectations, half-normal levels of MTP reduced plasma apoB-100 levels to the same extent (~25–35%) at each level of apoB synthesis. In addition, apoB secretion from primary hepatocytes was reduced to a comparable extent at each level of apoB synthesis. Thus, these results indicate that the concentration of MTP within the endoplasmic reticulum, rather than the MTP:apoB ratio, is the critical determinant of lipoprotein secretion. Finally, heterozygosity for an apoB knockout mutation was found to lower plasma apoB-100 levels more than heterozygosity for an MTP knockout allele. Consistent with that result, hepatic triglyceride accumulation was greater in heterozygous apoB knockout mice than in heterozygous MTP knockout mice. *Cre/loxP* tissue-specific recombination techniques were also used to generate liver-specific *Mtp* knockout mice. Inactivation of the *Mtp* gene in the liver caused a striking reduction in very low density lipoprotein (VLDL) triglycerides and large reductions in both VLDL/low density lipoproteins (LDL) and high density lipoprotein cholesterol levels. Histologic studies in liver-specific knockout mice revealed moderate hepatic steatosis. Currently being tested is the hypothesis that accumulation of triglycerides in the liver renders the liver more susceptible to injury by a second insult (e.g., lipopolysaccharide).

Human apo B (apolipoprotein B) Transgene mice show apo B locus may have a causative role male infertility The fertility of apoB (apolipoprotein B) (+/-) mice was recorded during the course of backcrossing (to C57BL/6J mice) and test mating. No apparent fertility problem was

Attorney's Docket No.: 14174-072W01

observed in female apoB (+/-) and wild-type female mice, as was documented by the presence of vaginal plugs in female mice. Although apoB (+/-) mice mated normally, only 40% of the animals from the second backcross generation produced any offspring within the 4-month test period. Of the animals that produced progeny, litters resulted from < 50% of documented matings. In contrast, all wild-type mice (6/6—i.e., 100%) tested were fertile. These data suggest genetic influence on the infertility phenotype, as a small number of male heterozygotes were not sterile. Fertilization *in vivo* was dramatically impaired in male apoB (+/-) mice. 74% of eggs examined were fertilized by the sperm from wild-type mice, whereas only 3% of eggs examined were fertilized by the sperm from apoB (+/-) mice. The sperm counts of apoB (+/-) mice were mildly but significantly reduced compared with controls. However, the percentage of motile sperm was markedly reduced in the apoB (+/-) animals compared with that of the wild-type controls. Of the sperm from apoB (+/-) mice, 20% (i.e., 4.9% of the initial 20% motile sperm) remained motile after 6 hr of incubation, whereas 45% (i.e., 33.6% of the initial 69.5%) of the motile sperm retained motility in controls after this time. *In vitro* fertilization yielded no fertilized eggs in three attempts with apo B (+/-) mice, while wild-type controls showed a fertilization rate of 53%. However, sperm from apoB (+/-) mice fertilized 84% of eggs once the zona pellucida had been removed. Numerous sperm from apoB (+/-) mice were seen binding to zona-intact eggs. However, these sperm lost their motility when observed 4–6 hours after binding, showing that sperm from apoB (+/-) mice were unable to penetrate the zona pellucida but that the interaction between sperm and egg was probably not direct. Sperm binding to zona-free oocytes was abnormal. In the apoB (+/-) mice, sperm binding did not attenuate, even after pronuclei had clearly formed, suggesting that apoB deficiency results in abnormal surface interaction between the sperm and egg.

Knockout of the mouse apoB gene resulted in embryonic lethality in homozygotes, protection against diet-induced hypercholesterolemia in heterozygotes, and developmental abnormalities in mice.

Model of insulin resistance, dyslipidemia & overexpression of human apoB It was shown that the livers of apoB mice assemble and secrete increased numbers of VLDL particles.

Example 2. Treatment of Diabetes Type-2 with iRNA

Introduction The regulation of hepatic gluconeogenesis is an important process in the adjustment of the blood glucose level. Pathological changes in the glucose production of the liver are a central characteristic in type-2-diabetes. For example, the fasting hyperglycemia
5 observed in patients with type-2-diabetes reflects the lack of inhibition of hepatic gluconeogenesis and glycogenolysis due to the underlying insulin resistance in this disease. Extreme conditions of insulin resistance can be observed for example in mice with a liver-specific insulin receptor knockout ('LIRKO'). These mice have an increased expression of the two rate-limiting gluconeogenic enzymes, phosphoenolpyruvate carboxykinase (PEPCK) and the
10 glucose-6-phosphatase catalytic subunit (G6Pase). Insulin is known to repress both PEPCK and G6Pase gene expression at the transcriptional level and the signal transduction involved in the regulation of G6Pase and PEPCK gene expression by insulin is only partly understood. While PEPCK is involved in a very early step of hepatic gluconeogenesis (synthesis of phosphoenolpyruvate from oxaloacetate), G6Pase catalyzes the terminal step of both,
15 gluconeogenesis and glycogenolysis, the cleavage of glucose-6-phosphate into phosphate and free glucose, which is then delivered into the blood stream.

The pharmacological intervention in the regulation of expression of PEPCK and G6Pase can be used for the treatment of the metabolic aberrations associated with diabetes. Hepatic glucose production can be reduced by an iRNA-based reduction of PEPCK and G6Pase
20 enzymatic activity in subjects with type-2-diabetes.

Targets for iRNA**Glucose-6-phosphatase (G6Pase)**

G6Pase mRNA is expressed principally in liver and kidney, and in lower amounts in the
25 small intestine. Membrane-bound G6Pase is associated with the endoplasmic reticulum. Low activities have been detected in skeletal muscle and in astrocytes as well.

G6Pase catalyzes the terminal step in gluconeogenesis and glycogenolysis. The activity of the enzyme is several fold higher in diabetic animals and probably in diabetic humans. Starvation and diabetes cause a 2-3-fold increase in G6Pase activity in the liver and a 2-4-fold
30 increase in G6Pase mRNA.

Attorney's Docket No.: 14174-072W01

Phosphoenolpyruvate carboxykinase (PEPCK)

Overexpression of PEPCK in mice results in symptoms of type-2-diabetes mellitus. PEPCK overexpression results in a metabolic pattern that increases G6Pase mRNA and results in a selective decrease in insulin receptor substrate (IRS)-2 protein, decreased phosphatidylinositol 3-kinase activity, and reduced ability of insulin to suppress gluconeogenic gene expression.

Table 5. Other targets to inhibit hepatic glucose production

Target	Comment
FKHR	good evidence for antidiabetic phenotype (Nakae <i>et al.</i> , <i>Nat Genetics</i> 32:245(2002))
Glucagon	
Glucagon receptor	
Glycogen phosphorylase	
PGC-1 (PPAR-Gamma Coactivator)	regulates the cAMP response (and probably the PKB/FKHR-regulation) on PEPCK/G6Pase
Fructose-1,6-bisphosphatase	
Glucose-6-phosphate translocator	
Glucokinase inhibitory regulatory protein	

Materials and Methods

Animals: BKS.Cg-m +/- Lepr db mice, which contain a point mutation in the leptin receptor gene are used to examine the efficacy of iRNA for the targets listed above.

BKS.Cg-m +/- Lepr db are available from the Jackson Laboratory (Stock Number 000642). These animals are obese at 3-4 weeks after birth, show elevation of plasma insulin at 10 to 14 days, elevation of blood sugar at 4 to 8 weeks, and uncontrolled rise in blood sugar. Exogenous insulin fails to control blood glucose levels and gluconeogenic activity increases.

The following numbers of male animals (age>12 weeks) could be tested with the following iRNAs:

- PEPCK, 2 sequences, 5 animals per sequence
- G6Pase, 2 sequences, 5 animals per sequence
- 1 nonspecific sequence, 5 animals
- 1 control group (only injected, no siRNA), 5 animals
- 1 control group (not injected, no siRNA), 5 animals

Attorney's Docket No.: 14174-072W01

Reagents: Necessary reagents would ideally include a Glucometer Elite XL (Bayer, Pittsburgh, PA) for glucose quantification, and an Insulin Radioimmunoassay (RIA) kit (Amersham, Piscataway, NJ) for insulin quantitation.

5 Assays:

G6P enzyme assays and PEPCK enzyme assays are used to measure the activity of the enzymes. Northern blotting is used to detect levels of G6Pase and PEPCK mRNA. Antibody-based techniques (e.g., immunoblotting, immunofluorescence) are used to detect levels of G6Pase and PEPCK protein. Glycogen staining is used to detect levels of glycogen in the liver. Histological
10 analysis is performed to analyze tissues.

Gene information:

G6Pase GenBank® No.: NM_008061, Mus musculus glucose-6-phosphatase, catalytic (G6pc), mRNA 1..2259, ORF 83..1156;

15 GenBank® No: U00445, Mus musculus glucose-6-phosphatase mRNA, complete cds 1..2259, ORF 83..1156

GenBank® No: BC013448

PEPCK

GenBank® No: NM_011044, Mus musculus phosphoenolpyruvate carboxykinase 1, cytosolic (Pck1), mRNA.1..2618, ORF 141..2009
20

GenBank® No: AF009605.1

Administration of iRNA:

iRNA corresponding to the genes described above could be administered to mice with
25 hydrodynamic injection. One control group of animals would be treated with Metformin as a positive control for reduction in hepatic glucose levels.

Experimental Protocol

Mice could be housed in a facility in which there is light from 7:00 AM to 7:00 PM.
30 Mice would be fed *ad libitum* from 7:00 PM to 7:00 AM and fast from 7:00 AM to 7:00 PM.

Attorney's Docket No.: 14174-072W01

Day 0: 7:00 PM: Approximately 100 μ l blood would be drawn from the tail. Serum could be isolated to measure glucose, insulin, HbA1c (EDTA-blood), glucagon, FFAs, lactate, corticosterone, serum triglycerides.

5 **Day 1-7:** Blood glucose could be measured daily at 8:00 AM and 6:00 PM (approx. 3-5 μ l; measured with a Haemoglucometer)

Day 8: Blood glucose could be measured daily at 8:00 AM and 6:00 PM. iRNA would be injected between 10:00 AM and 2:00 PM

10 **Day 9-20:** Blood glucose could be measured daily at 8:00 AM and 6:00 PM.

Day 21: Mice could be sacrificed after 10 hours of fasting.

Blood would be isolated. Glucose, insulin, HbA1c (EDTA-blood), glucagon, FFAs, lactate, corticosterone, serum triglycerides would be measured. Liver tissue would be isolated for histology, protein assays, RNA assays, glycogen quantitation, and enzyme assays.

15

Example 3: Inhibition of Glucose-6-Phosphatase iRNA *in vivo*

iRNA targeted to the Glucose-6-Phosphatase (G6P) gene was used to examine the effects of inhibition of G6P expression on glucose metabolism *in vivo*.

20 Female mice, 10 weeks of age, strain BKS.Cg-m +/- Lepr db (The Jackson Laboratory) were used for *in vivo* analysis of enzymes of the hepatic glucose production. Mice were housed under conditions where it was light from 6:30 am to 6:30 pm. Mice were fed (ad libidum) during the night period and fasted during the day period.

25 On day 1, approximately 100 μ l of blood was collected from test animals by puncturing the retroorbital plexus. On days 1-7, blood glucose was measured in blood obtained from tail veins (approximately 3-5 μ l) using a Glucometer (Elite XL, Bayer). Blood glucose was sampled daily at 8 am and 6 pm.

On day 7 at approximately 2pm, GL3 plasmid (10 μ g) and siRNAs (100 μ g G6Pase specific, Renilla nonspecific or no siRNA control) were delivered to animals using hydrodynamic coinjection.

Attorney's Docket No.: 14174-072W01

On day 8, GL3 expression was analyzed by injection of luciferin (3 mg) after anaesthesia with avertin and imaging. This was done to control for successful hydrodynamic delivery.

On days 8-10, blood glucose was measured in blood obtained from tail veins (approximately 3-5 ml) using a Glucometer (Elite XL, Bayer).

5 On day 10, mice were sacrificed after 10 hours of fasting. Blood and liver were isolated from sacrificed animals.

Table 6 lists blood glucose levels (mg/dl) for mice injected with GL3 plasmid and G6Pase iRNA (G6P4), Renilla nonspecific iRNA (RL), or no iRNA (no). Days on which nucleic acids were injected are shaded.

10 Table 6. Blood glucose levels in mice

plasmid siRNA	GL3 G6P4 mouse 03	GL3 G6P4 mouse 04	GL3 G6P4 mouse 05	GL3 no mouse 07	GL3 G6P4 mouse 09	GL3 RL mouse 14	GL3 RL mouse 15	GL3 no mouse 17
day	BG	BG	BG	BG	BG	BG	BG	BG
1	512	250	337	241	196	275	538	437
2	555	437	339	556	406	315	524	386
3	483	446	356	567	283	491	900	459
4	579	543	552	423	404	457	548	375
5	800	501	800	277	196	441	533	433
6	484	600	408	454	461	412	490	301
7	214	209	245	260	284	600	425	injection
8	500	566	246	521	277	600	578	404
9	369	448	438			596	600	459
10	369	600	446					injection
average day 1 to 8	532	463	465	420	325	399	539	396

15 Table 7 lists average blood glucose levels (mg/dl) on days 1-6 or day 7 for mice injected with GL3 plasmid and G6Pase iRNA (G6P4), Renilla nonspecific iRNA (RL), no iRNA (no), or for mice that were not injected, or for which injection failed.

20 Table 7. Average blood glucose levels

	G6P4	RL	no	RL and no (combined)
	mouse 03,04,05,09	mouse 14, 15	mouse 07, 17	mouse 14, 15, 07, 17
average (d1-6)	446	469	409	439
stddev (d1-6)	124	96	101	101
average (d7)	230	547	369	458
stddev (d7)	27			122

FIGs. 6A, 6B, and 6C show graphs depicting blood glucose levels of animals injected with control or no siRNA, G6Pase RNA, or non-injected mice (respectively) at days 1-6 and day 7. FIG. 7 contains a graph of average blood glucose levels for mice injected with G6Pase RNA (solid line) and mice injected with, Renilla nonspecific iRNA (RL) or no iRNA (no) (dashed line).

Table 8 lists average blood glucose levels for mice injected with G6Pase iRNA or Renilla nonspecific iRNA (RL) and no iRNA.

Table 8. Average blood glucose levels

iRNA	G6P4		RL and no (combined)		
	mouse 03,04,05,09		mouse 14, 15, 07, 17		
	average	stddev	average	stddev	
day					
1	374	176	373		139
2	435	90	445		114
3	392	90	529		65
4	520	79	451		73
5	475	190	420		106
6	483	82	414		82
7	230	27	458		122
8	422	187	525		87
9	412	54	532		71
10	472	118			

Example 4: Selected Palindromic Sequences

- Tables 9-14 below provide selected palindromic sequences from the following genes: human ApoB, human glucose-6-phosphatase, rat glucose-6-phosphatase, β -catenin, and hepatitis C virus (HCV).

Table 9. Selected palindromic sequences from human ApoB

	Source	Start Index	End Index		Match	Start Index	End Index	#B
SEQ ID NO: 1	ggccattccagaaggggaag	509	528	SEQ ID NO: 1004	cttcctgtctglaatggcc	5795	5814	19
SEQ ID NO: 2	tgcattctcgagagtcca	4099	4118	SEQ ID NO: 1005	tggaaactctdccaatggca	10876	10895	18
SEQ ID NO: 3	catgtcaaacactttgtta	7056	7075	SEQ ID NO: 1006	taacaaattcttgacatg	7358	7377	18
SEQ ID NO: 4	tttgtataaactctattg	7068	7087	SEQ ID NO: 1007	caataagatcaatagcaaa	8990	9009	18
SEQ ID NO: 5	tctgaaaagggtcatgga	8880	8899	SEQ ID NO: 1008	tccatgtcccatattacaga	11356	11375	18
SEQ ID NO: 6	cagctctgttcagggtcca	10900	10919	SEQ ID NO: 1009	tggacctgcaccaaaagctg	13952	13971	18
SEQ ID NO: 7	ggagggtcccatctctgc	356	375	SEQ ID NO: 1010	gcagccctgggaanaactcc	6447	6466	17
SEQ ID NO: 8	ctgtttgaagactctcca	1081	1100	SEQ ID NO: 1011	tggagggtagtataacag	10327	10346	17
SEQ ID NO: 9	agtggtctgaaactgtgtca	1297	1316	SEQ ID NO: 1012	tgcagagcttttgcact	13508	13527	17
SEQ ID NO: 10	ccaaaatagaagggaatct	2068	2087	SEQ ID NO: 1013	agatctcttgccttttg	4000	4019	17
SEQ ID NO: 11	tgaagagaagatgaattt	3620	3639	SEQ ID NO: 1014	aaattctctttcttttca	9212	9231	17
SEQ ID NO: 12	agtggtgtgcaacaccagca	4230	4249	SEQ ID NO: 1015	tgtagtgtgagccaacact	10649	10668	17
SEQ ID NO: 13	aaggctccacaagtcatca	5950	5969	SEQ ID NO: 1016	tgatgatcatctggaacctt	10724	10743	17
SEQ ID NO: 14	gtcagccagggttatagca	7725	7744	SEQ ID NO: 1017	tgtctagaaccttactgac	7781	7800	17
SEQ ID NO: 15	tgatacttggaacttgaa	10727	10746	SEQ ID NO: 1018	ttcactgttctgaaatca	7863	7882	17
SEQ ID NO: 16	gtcaagttgagcaatttct	13423	13442	SEQ ID NO: 1019	agaaaaggcacaccttgac	11072	11091	17
SEQ ID NO: 17	atccagatggaaaaggga	13480	13499	SEQ ID NO: 1020	ttccaatttccctgtgat	3680	3699	17
SEQ ID NO: 18	atttgtttgtcaagaagt	4543	4562	SEQ ID NO: 1021	acttcagagaaatacaaat	11401	11420	46
SEQ ID NO: 19	ctggaanaagtcagctcgg	204	223	SEQ ID NO: 1022	ccagactctcgtttaccag	8235	8254	26
SEQ ID NO: 20	accaggagggttctcttca	1729	1748	SEQ ID NO: 1023	tgaagtgtactctcgtgt	5089	5108	26
SEQ ID NO: 21	aaagaagttctgaagaagt	1956	1975	SEQ ID NO: 1024	attccatcacaaalccttt	9661	9680	26
SEQ ID NO: 22	gtcatcagctatggctcca	3570	3589	SEQ ID NO: 1025	tggatctaaagcagtagc	11623	11642	26
SEQ ID NO: 23	atcaatatgatcaatttg	6414	6433	SEQ ID NO: 1026	caaagaagtcaagattgat	4553	4572	26
SEQ ID NO: 24	gaattacttttaaacat	7326	7345	SEQ ID NO: 1027	atgtgttaacaaaattatc	11494	11513	26
SEQ ID NO: 25	cgaggccgcgtctgtggc	130	149	SEQ ID NO: 1028	gccagaagtgtgactctcg	3507	3526	16
SEQ ID NO: 26	acaactatgaggctgagag	271	290	SEQ ID NO: 1029	ctctgagcaacaaattgtt	10309	10328	16

SEQ ID NO:	27	gctgagagttccagtgagg	282	301	SEQ ID NO:	1030	ctcatggcfaatgtcagc	10885	10904	16
SEQ ID NO:	28	gaagaaaaccaagaactc	448	467	SEQ ID NO:	1031	gagtcattgaggttctca	4929	4948	16
SEQ ID NO:	29	cctacttacctcctgaaca	558	577	SEQ ID NO:	1032	gttcataagggaggtagg	12766	12785	16
SEQ ID NO:	30	ctacttacctcctgaacat	559	578	SEQ ID NO:	1033	atgttcataagggaggttag	12765	12784	16
SEQ ID NO:	31	gagacagaagaagccaagc	615	634	SEQ ID NO:	1034	gcttggttttgccagltctc	2459	2478	16
SEQ ID NO:	32	cactcactttacogtcaag	671	690	SEQ ID NO:	1035	cttgaacacaaaagtcagtg	6000	6019	16
SEQ ID NO:	33	ctgatcagcagcagccagt	822	841	SEQ ID NO:	1036	actgggaagtgcitacag	5237	5256	16
SEQ ID NO:	34	actggacgctaagaggaag	854	873	SEQ ID NO:	1037	cttcccaaaagagaccagt	2890	2909	16

SEQ ID NO:35	agaggagcatggtgcaga	865	884	SEQ ID NO:1038	tcggcattactttctct	5921	5940	16
SEQ ID NO:36	tgaagactctccaggaact	1087	1106	SEQ ID NO:1039	agtgaaggagactattca	7216	7235	16
SEQ ID NO:37	ctcagcagaaaatccag	1121	1140	SEQ ID NO:1040	ctggttactgacgagag	1181	1180	16
SEQ ID NO:38	atgaagcagtcacatctct	1189	1208	SEQ ID NO:1041	agagctgccagctctcat	10016	10035	16
SEQ ID NO:39	ttgccacagctgattgagg	1209	1228	SEQ ID NO:1042	cctcctacaggtgggcaa	4222	4241	16
SEQ ID NO:40	agctgattgagggtccag	1216	1235	SEQ ID NO:1043	ctggttccacatgcagct	11847	11866	16
SEQ ID NO:41	tgctccactcacatctcc	1278	1297	SEQ ID NO:1044	ggaggcttaagtgcagca	7601	7620	16
SEQ ID NO:42	tgaacgtgtgcagccaa	1303	1322	SEQ ID NO:1045	ttgggagagacaagttca	6500	6519	16
SEQ ID NO:43	gacattgctaattacctga	1503	1522	SEQ ID NO:1046	tcagaagctaaagcaatgc	7232	7251	16
SEQ ID NO:44	ttctctcagactttct	1738	1757	SEQ ID NO:1047	aggagagccaattagaa	8498	8517	16
SEQ ID NO:45	ccaatatctgaactcaga	1903	1922	SEQ ID NO:1048	tcgaattcattcaattgg	6485	6504	16
SEQ ID NO:46	aaagttagtgaaagaagt	1946	1965	SEQ ID NO:1049	aactacctcactgcttt	2132	2151	16
SEQ ID NO:47	aagttagtgaaagaagtc	1947	1966	SEQ ID NO:1050	gaacctctggcattact	5916	5935	16
SEQ ID NO:48	aaagaagttcgaagaat	1956	1975	SEQ ID NO:1051	attctctggtaactacttt	5482	5501	16
SEQ ID NO:49	tttgctataccaagaatg	2322	2341	SEQ ID NO:1052	catcttaggcactgacaaa	4997	5016	16
SEQ ID NO:50	tgttgagaagctgalttaa	2381	2400	SEQ ID NO:1053	tttagccatcggtcaaca	5700	5719	16
SEQ ID NO:51	caggaagggtcaagaat	2561	2580	SEQ ID NO:1054	attcctttaacaattctg	9492	9511	16
SEQ ID NO:52	aggaagggtcaagaagt	2562	2581	SEQ ID NO:1055	cattcctttaacaattct	9491	9510	16
SEQ ID NO:53	gaagggtcaaagaatgac	2564	2583	SEQ ID NO:1056	gtcagctctcaggctcttc	7914	7933	16
SEQ ID NO:54	caaagaatgactttttct	2572	2591	SEQ ID NO:1057	agaaggatggcatttttg	14000	14019	16
SEQ ID NO:55	catggagaatgacttgaa	2603	2622	SEQ ID NO:1058	ttcagagccaagtcacat	7119	7138	16
SEQ ID NO:56	ggagccaaggtcgagtaa	2679	2698	SEQ ID NO:1059	ttactccaacgcagctcc	3050	3069	16
SEQ ID NO:57	tcattctctcccaagag	2884	2903	SEQ ID NO:1060	ctctctgggcatctatga	5139	5158	16
SEQ ID NO:58	acctatgagctccagagag	3165	3184	SEQ ID NO:1061	ctctcaagaccacagaggt	12976	12995	16
SEQ ID NO:59	gggcaaacgtcttcagaa	3365	3384	SEQ ID NO:1062	tcgaaagacaacgtgcc	12317	12336	16
SEQ ID NO:60	acctgagcattcagaaca	3387	3406	SEQ ID NO:1063	tggtgctaaggttcagggt	5675	5694	16
SEQ ID NO:61	atgggcgacctaatgttg	3429	3448	SEQ ID NO:1064	cacaattagtttcccat	8941	8960	16
SEQ ID NO:62	gatgaagagaagattgaat	3618	3637	SEQ ID NO:1065	attcagcttccccacatc	8330	8349	16

				NO:									
SEQ ID NO: 63	caatgataaccacaaaaa	3656	3675	SEQ ID NO:	1066	tttttggaatgccattg	8643	8662	16				
SEQ ID NO: 64	gtagataccaaaaaatga	3660	3679	SEQ ID NO:	1067	tcatgtgatgggtctctac	4371	4390	16				
SEQ ID NO: 65	gcttcagttcatttggact	4509	4528	SEQ ID NO:	1068	agtcaagaaggacttaagc	5304	5323	16				
SEQ ID NO: 66	ttgtttgtcaaaagaagtc	4544	4563	SEQ ID NO:	1069	gacttcagagaaatacaaa	11400	11419	16				
SEQ ID NO: 67	ttgtttgtcaaaagaagtc	4545	4564	SEQ ID NO:	1070	tgacttcagagaaatacaa	11399	11418	16				
SEQ ID NO: 68	tggcaatgggaactcgtct	5846	5865	SEQ ID NO:	1071	agcgagaatcacccgtcca	8219	8238	16				
SEQ ID NO: 69	aacctctggcattacttt	5917	5936	SEQ ID NO:	1072	aaaggagatgtcaaggggt	10599	10618	16				
SEQ ID NO: 70	catttactttctctcatga	5926	5945	SEQ ID NO:	1073	tcattgaaagaataaatg	7026	7045	16				
SEQ ID NO: 71	aaagtcatgtccctgctta	6009	6028	SEQ ID NO:	1074	taagaaccttactgacttt	7784	7803	16				
SEQ ID NO: 72	tccatttttgagacctt	6322	6341	SEQ ID NO:	1075	aaggacttcaggaatggga	12004	12023	16				
SEQ ID NO: 73	catcaatattgatcaatt	6413	6432	SEQ ID NO:	1076	aaattaaaaagtcttgatg	6732	6751	16				
SEQ ID NO: 74	taaagatagattatgatta	6665	6684	SEQ ID NO:	1077	taaaccaaaactgggtta	9019	9038	16				
SEQ ID NO: 75	tattgatgaatcattgaa	6713	6732	SEQ ID NO:	1078	ttcaagacttaaaaaata	8007	8026	16				
SEQ ID NO: 76	atgatctacattgtttat	6790	6809	SEQ ID NO:	1079	ataaagaaataaagtcac	7380	7399	16				
SEQ ID NO: 77	agagacacatacagaatat	6919	6938	SEQ ID NO:	1080	atataattgcagtgcctct	13382	13401	16				
SEQ ID NO: 78	gacacatacagaatataga	6922	6941	SEQ ID NO:	1081	tctaaattcagttctgttc	11327	11346	16				
SEQ ID NO: 79	agcatgtcaaacactttgt	7054	7073	SEQ ID NO:	1082	acaagtgcagtgccctgct	6007	6026	16				
SEQ ID NO: 80	tttttagaggaaaccaagg	7515	7534	SEQ ID NO:	1083	cctttgtgacacacaaaa	11230	11249	16				
SEQ ID NO: 81	tttttagaggaaaccaagg	7516	7535	SEQ ID NO:	1084	gcctttgtgacacacaaaa	11229	11248	16				
SEQ ID NO: 82	ggaagatagacttctgaa	9307	9326	SEQ ID NO:	1085	ttcagaataactgttttcc	12824	12843	16				
SEQ ID NO: 83	cactgtttctgagtcacag	9334	9353	SEQ ID NO:	1086	ctgggacctaccaagagtg	12523	12542	16				
SEQ ID NO: 84	cacaatcctttggctgtg	9668	9687	SEQ ID NO:	1087	cacattcaaggaattgtg	10063	10082	16				
SEQ ID NO: 85	ttctggatacacgtttcc	9853	9872	SEQ ID NO:	1088	ggaactgttgactcaggaa	12569	12588	16				
SEQ ID NO: 86	gaaatctcaagtttctct	10042	10061	SEQ ID NO:	1089	agagccaggtcgagctttc	11044	11063	16				
SEQ ID NO: 87	tttctcatctcatctgt	10210	10229	SEQ ID NO:	1090	acagctgaagagatgaaa	13055	13074	16				
SEQ ID NO: 88	ttaccgctaaggagcag	10521	10540	SEQ ID NO:	1091	ctgcacgctttgaggtaga	11761	11780	16				
SEQ ID NO: 89	ctaccgctaaggagcag	10522	10541	SEQ ID NO:	1092	actgcacgctttgaggtag	11760	11779	16				
SEQ ID NO: 90	agggcctcttttaccacaa	10831	10850	SEQ ID NO:	1093	ttggccagggaagtgccct	10957	10976	16				

				NO:									
SEQ ID NO: 91	ttctccatccctgtaaaag	11285	11284	SEQ ID NO:	1094	cttttccaccaacggagaa	10838	10857	1	6			
SEQ ID NO: 92	gaaaaacaaagcagattat	11816	11835	SEQ ID NO:	1095	ataaactgaagattttc	13600	13619	1	6			
SEQ ID NO: 93	actcactcattgattttct	12682	12701	SEQ ID NO:	1096	agaaaaacaggatctgagt	14027	14046	1	6			
SEQ ID NO: 94	taaaactaatagatgaatc	12890	12909	SEQ ID NO:	1097	gatlaccaccagcagttta	13578	13597	1	6			
SEQ ID NO: 95	caaacgagctcaggaag	13200	13219	SEQ ID NO:	1098	ctcgtgaagaataatttg	13260	13279	1	6			
SEQ ID NO: 96	tgaataatgctcagtggt	2366	2385	SEQ ID NO:	1099	aacacttactgaattcca	10662	10681	3	5			
SEQ ID NO: 97	gatttgaatccaaagaag	2400	2419	SEQ ID NO:	1100	cttcagagaaatacaaatc	11402	11421	3	5			
SEQ ID NO: 98	atttgaatccaaagaagt	2401	2420	SEQ ID NO:	1101	acttcagagaaatacaaat	11401	11420	3	5			
SEQ ID NO: 99	atcaacagcgcctctttg	990	1009	SEQ ID NO:	1102	caaagaagtcaagattgat	4553	4572	2	5			
SEQ ID NO: 100	tgtttgaagactctccag	1082	1101	SEQ ID NO:	1103	ctggaaagttaaacaaca	6955	6974	2	5			
SEQ ID NO: 101	ccctctgatagatgttgt	1324	1343	SEQ ID NO:	1104	accaagactggcaccaggg	13961	13980	2	5			
SEQ ID NO: 102	tgagcaagtgaagaacttt	1868	1887	SEQ ID NO:	1105	aaagccattcagctctcca	12963	12982	2	5			
SEQ ID NO: 103	atttgaatccaaagaagt	2401	2420	SEQ ID NO:	1106	actttctaaacttgaat	9055	9074	2	5			
SEQ ID NO: 104	atccaaagaagtcocggaa	2408	2427	SEQ ID NO:	1107	ttccggggaaacccgggat	12721	12740	2	5			
SEQ ID NO: 105	agagcctacctcgcctct	2430	2449	SEQ ID NO:	1108	agatggtacgttagcctct	11921	11940	2	5			
SEQ ID NO: 106	aatgccttgaactcccca	2610	2629	SEQ ID NO:	1109	tgggaactacaatttcatt	7012	7031	2	5			
SEQ ID NO: 107	gaagtccaattccggatt	3297	3316	SEQ ID NO:	1110	aatctcaattattcttc	13815	13834	2	5			
SEQ ID NO: 108	tgaagcagaagccagaag	3496	3515	SEQ ID NO:	1111	cttcaggttcacatctgca	11376	11395	2	5			
SEQ ID NO: 109	gaagagaagattgaatttg	3621	3640	SEQ ID NO:	1112	caaaaactactgtctcttc	10459	10478	2	5			
SEQ ID NO: 110	atgctaaaggcacatag	4597	4616	SEQ ID NO:	1113	ccatatgaagatcaagcat	12656	12675	2	5			
SEQ ID NO: 111	tcctcacctccacctctg	4737	4756	SEQ ID NO:	1114	cagattctcagatgagggg	8912	8931	2	5			
SEQ ID NO: 112	attlacagctctgacaagt	5427	5446	SEQ ID NO:	1115	actttctaaacttgaat	9055	9074	2	5			
SEQ ID NO: 113	aggagcctacaaaaataat	5594	5613	SEQ ID NO:	1116	attatgttgaacagctcct	11830	11849	2	5			
SEQ ID NO: 114	aaagtgaagcacatcaat	6401	6420	SEQ ID NO:	1117	attgttgcctcatctcctt	10194	10213	2	5			
SEQ ID NO: 115	ctgctggaacaaacagaaa	9418	9437	SEQ ID NO:	1118	ttctgattaccaccagcag	13574	13593	2	5			
SEQ ID NO: 116	tgaaggaattcttgaaaa	9582	9601	SEQ ID NO:	1119	ttttaaagaaatactcaa	13805	13824	2	5			
SEQ ID NO: 117	gaagtaaaagaaaatttg	10743	10762	SEQ ID NO:	1120	caaaaactactgtctcttc	10459	10478	2	5			
SEQ ID NO: 118	tgaagaagatgcgaattt	11984	12003	SEQ ID NO:	1121	aaatgtcagctctgttcca	10894	10913	2	5			

SEQ ID NO: 119	aggatctgagttatttgc	14035	14054	SEQ ID NO: 1122	gcaagtcagcccattct	10920	10939	25
SEQ ID NO: 120	gtgccttctcggttgctg	18	37	SEQ ID NO: 1123	cagccattgacatgagcac	5740	5759	15
SEQ ID NO: 121	ggcgtgctgcgtgctg	146	165	SEQ ID NO: 1124	cagctccacagactccgcc	3062	3081	15
SEQ ID NO: 122	ctgcgtgctgctgctg	154	173	SEQ ID NO: 1125	agcagaaggtggaagcag	3224	3243	15
SEQ ID NO: 123	gctgctggcggcgccagg	170	189	SEQ ID NO: 1126	cctgattccacatgcagc	11846	11865	15
SEQ ID NO: 124	aagaggaaatgctggaaaa	193	212	SEQ ID NO: 1127	ttttctcactacatctt	2584	2603	15
SEQ ID NO: 125	ctggaaaatgtcagctgg	204	223	SEQ ID NO: 1128	ccagactccacatccac	3915	3934	15
SEQ ID NO: 126	tggagtccctgggactgct	296	315	SEQ ID NO: 1129	agcatgctgatttctcca	9945	9964	15
SEQ ID NO: 127	ggagtcctgggactgctg	297	316	SEQ ID NO: 1130	cagcatgctgatttctcc	9944	9963	15
SEQ ID NO: 128	tgggactgctgattcaaga	305	324	SEQ ID NO: 1131	tcttcactcactgaccca	2042	2061	15
SEQ ID NO: 129	ctgctgattcaagaagtgc	310	329	SEQ ID NO: 1132	gcacacctgacattgcag	11079	11098	15
SEQ ID NO: 130	tggccaccagatcaactgc	326	345	SEQ ID NO: 1133	gcaggctgaactggtggca	2717	2736	15
SEQ ID NO: 131	gccaccagatcaactgca	327	346	SEQ ID NO: 1134	tgcaggctgaactggtggc	2716	2735	15
SEQ ID NO: 132	tgcagggtgagctggagg	342	361	SEQ ID NO: 1135	ctccacctctgatctgca	4744	4763	15
SEQ ID NO: 133	caagggtgagctggaggtt	344	363	SEQ ID NO: 1136	aacccctacatgaagcttg	13755	13774	15
SEQ ID NO: 134	ctctgcagctcatctga	369	388	SEQ ID NO: 1137	tcaggaagcttcaagag	13211	13230	15
SEQ ID NO: 135	cagctcatctgaagacc	374	393	SEQ ID NO: 1138	ggcttgagttaatgctg	4977	4996	15
SEQ ID NO: 136	gcttcatctgaagaccag	376	395	SEQ ID NO: 1139	ctggaagctgaagaggaac	855	874	15
SEQ ID NO: 137	tcatctgaagaccagcca	379	398	SEQ ID NO: 1140	tggcatggcattatgatga	3604	3623	15
SEQ ID NO: 138	gaaaaccaagaactctgag	452	471	SEQ ID NO: 1141	ctcaaccttaatgatctt	8286	8305	15
SEQ ID NO: 139	agaactctgaggatttgc	460	479	SEQ ID NO: 1142	gcaagctacagatttct	8377	8396	15
SEQ ID NO: 140	ctcgaggagtttgctgcag	465	484	SEQ ID NO: 1143	ctgcaggggatccccagaga	2526	2545	15
SEQ ID NO: 141	tttgctgcagccatgtcca	474	493	SEQ ID NO: 1144	tggagtgctcagggcaaa	10372	10391	15
SEQ ID NO: 142	caagaggggcatcatttct	578	597	SEQ ID NO: 1145	agaataaatgacgttctg	7035	7054	15
SEQ ID NO: 143	tcactttaccgtcaagagc	674	693	SEQ ID NO: 1146	cgctacactatcatgtga	4360	4379	15
SEQ ID NO: 144	tttaccgtcaagacgagga	678	697	SEQ ID NO: 1147	tcttgacatgttgataaa	7366	7385	15
SEQ ID NO: 145	cactggacgtaagaggaa	853	872	SEQ ID NO: 1148	ttccagaagcagccagtg	12498	12517	15
SEQ ID NO: 146	aggaagcatgtggcagaag	867	886	SEQ ID NO: 1149	cttcatacacattaatcct	9988	10007	15

SEQ ID NO:	147	caaggagcaacacctcttc	893	912	SEQ ID NO:	1150	gaagtactagtcactctg	6835	6854	15
SEQ ID NO:	148	acagacttgaacttgaa	959	978	SEQ ID NO:	1151	ttcaattctcaatgcigt	10500	10519	15
SEQ ID NO:	149	tgatgaagcagtcacatct	1187	1206	SEQ ID NO:	1152	agatttgaggattccatca	7976	7995	15
SEQ ID NO:	150	agcagtcacatctctctg	1193	1212	SEQ ID NO:	1153	caaggagaaactgactgct	6524	6543	15
SEQ ID NO:	151	ccagcccatcactttaca	1231	1250	SEQ ID NO:	1154	tgtagtctcctgggtctg	5094	5113	15
SEQ ID NO:	152	ctccactcacatctctcag	1280	1299	SEQ ID NO:	1155	ctggagcttagtaatggag	8709	8728	15
SEQ ID NO:	153	catgccacccccctctga	1314	1333	SEQ ID NO:	1156	tcagatggagggaacacatg	8919	8938	15
SEQ ID NO:	154	gagagatctcaacatggc	1390	1409	SEQ ID NO:	1157	gccaccctggaactctctc	10869	10888	15
SEQ ID NO:	155	lcaacatggcgaggatca	1399	1418	SEQ ID NO:	1158	tgatcccaactctcattga	2965	2984	15
SEQ ID NO:	156	ccacctgtatgcctgag	1429	1448	SEQ ID NO:	1159	ctcagggatctgaaggtgg	8187	8206	15
SEQ ID NO:	157	gtcaacaactatcataaga	1455	1474	SEQ ID NO:	1160	tcttgatttaaatctgac	4979	4998	15
SEQ ID NO:	158	tgacattgctaattacct	1501	1520	SEQ ID NO:	1161	aggtatatcgaaagtcca	12799	12818	15
SEQ ID NO:	159	ggacattgctaattacctg	1502	1521	SEQ ID NO:	1162	caggatatatcgaaagtcc	12798	12817	15
SEQ ID NO:	160	ttctgctgggtatggaaa	1573	1592	SEQ ID NO:	1163	ttcacatgccaaggagaaa	6514	6533	15
SEQ ID NO:	161	ccagaactcaagttctcaa	1620	1639	SEQ ID NO:	1164	ttgaagtgtagtctcctg	5088	5107	15
SEQ ID NO:	162	agtcttcaatctgaaatg	1630	1649	SEQ ID NO:	1165	cattctgattggtggact	7757	7776	15
SEQ ID NO:	163	tgagcaagtgaagaacttt	1868	1887	SEQ ID NO:	1166	aaagtgccacttttactca	6183	6202	15
SEQ ID NO:	164	agcaagtgaagaactttgt	1870	1889	SEQ ID NO:	1167	acaagtgcagtgccctgct	6007	6026	15
SEQ ID NO:	165	tcgaaagaatctcaact	1964	1983	SEQ ID NO:	1168	aagtcataatggttcaga	12811	12830	15
SEQ ID NO:	166	actgtcatggacttcagaa	1986	2005	SEQ ID NO:	1169	ttctgaatatattgtcagt	13376	13395	15
SEQ ID NO:	167	actgacccagcctcagcc	2051	2070	SEQ ID NO:	1170	ggctcaccctgagagaagt	12391	12410	15
SEQ ID NO:	168	tcaataactacctctct	2096	2115	SEQ ID NO:	1171	aggaagatatgaagtga	4712	4731	15
SEQ ID NO:	169	actacccctcactgctttg	2133	2152	SEQ ID NO:	1172	caaatttggagggttagt	10319	10338	15
SEQ ID NO:	170	ttgattgtctcagctga	2149	2168	SEQ ID NO:	1173	tcagataagtaacaacaa	9392	9411	15
SEQ ID NO:	171	ttggaagctctttttggga	2211	2230	SEQ ID NO:	1174	tcogattcaagcttcaa	11577	11596	15
SEQ ID NO:	172	ggaagctctttttgggaag	2213	2232	SEQ ID NO:	1175	cttcagaaagctacctcc	7929	7948	15
SEQ ID NO:	173	ttttccagacagtgta	2238	2257	SEQ ID NO:	1176	tgaccttctcaagcaaaa	4876	4895	15
SEQ ID NO:	174	agacagtgtaacaaagct	2246	2265	SEQ ID NO:	1177	agcttggttttgcagct	2458	2477	15

SEQ ID NO: 175	cttggctataccaagat	2321	2340	SEQ ID NO: 1178	atctcgtgtctaggaaaag	5968	5987	15
SEQ ID NO: 176	caaagatgataaacatgag	2333	2352	SEQ ID NO: 1179	ctcaaggataacgtgttg	12609	12628	15
SEQ ID NO: 177	gatatgtlaaatggaataa	2355	2374	SEQ ID NO: 1180	ttatcttataattatctc	13079	13098	15
SEQ ID NO: 178	ggaataatgctcagtggtg	2367	2386	SEQ ID NO: 1181	caacacttactgaattcc	10661	10680	15
SEQ ID NO: 179	tttgaatccaagaagtc	2402	2421	SEQ ID NO: 1182	gacttcagagaaatacaaa	11400	11419	15
SEQ ID NO: 180	gatccccccagatgattgga	2534	2553	SEQ ID NO: 1183	tcaatttccctgtggatc	3681	3700	15
SEQ ID NO: 181	cagatgattggagaggta	2541	2560	SEQ ID NO: 1184	tgaccacacaacagctctg	5363	5382	15
SEQ ID NO: 182	agaatgactttttctca	2575	2594	SEQ ID NO: 1185	tgaagtcgggattctct	11015	11034	15
SEQ ID NO: 183	gaactccccactggagctg	2619	2638	SEQ ID NO: 1186	cagctcaaccgtacagttc	11861	11880	15
SEQ ID NO: 184	atatcttcatctggagta	2652	2671	SEQ ID NO: 1187	tgacttcagtcagaatat	11966	11985	15
SEQ ID NO: 185	gtcatgtctccggagcca	2667	2686	SEQ ID NO: 1188	tggccccgtttaaccatgac	5809	5828	15
SEQ ID NO: 186	gctgaagttaattctct	2873	2892	SEQ ID NO: 1189	aggaggctttaagttcagc	7600	7619	15
SEQ ID NO: 187	attccttccccaaagagac	2886	2905	SEQ ID NO: 1190	gtcttcttcccttggaat	10470	10489	15
SEQ ID NO: 188	ctcattgagaacaggcagt	2976	2995	SEQ ID NO: 1191	actgactgcacgctttgag	11756	11775	15
SEQ ID NO: 189	ttgagcagttattctgtag	3142	3161	SEQ ID NO: 1192	ctgagagaaagtgcttcaa	12399	12418	15
SEQ ID NO: 190	acctgtccagtgaaagtc	3285	3304	SEQ ID NO: 1193	ggacggtaactgtccagggt	12784	12803	15
SEQ ID NO: 191	ccagtgaggtccaattcc	3292	3311	SEQ ID NO: 1194	ggaaggcagagtttactgg	9148	9167	15
SEQ ID NO: 192	acattcagaacaagaaat	3384	3413	SEQ ID NO: 1195	atttcttaagctggagt	11167	11186	15
SEQ ID NO: 193	gaaaaatcaagggtgttat	3463	3482	SEQ ID NO: 1196	ataaactgcaagatttttc	13600	13619	15
SEQ ID NO: 194	aaatcaagggtgtatttc	3466	3485	SEQ ID NO: 1197	gaaacaatgcattagattt	9745	9764	15
SEQ ID NO: 195	tggcattatgatgaagaga	3609	3628	SEQ ID NO: 1198	tctcccggtataatgcca	11781	11800	15
SEQ ID NO: 196	aagagaagattgaattga	3622	3641	SEQ ID NO: 1199	tcaaaacctactgtctctt	10458	10477	15
SEQ ID NO: 197	aaatgactccaatttccc	3673	3692	SEQ ID NO: 1200	gggaactacaatttcattt	7013	7032	15
SEQ ID NO: 198	atgacttccaatttccctg	3675	3694	SEQ ID NO: 1201	caggctgatacagatcat	4917	4936	15
SEQ ID NO: 199	acttccaatttccctgtgg	3678	3697	SEQ ID NO: 1202	ccacgaaaaataggaagt	10360	10379	15
SEQ ID NO: 200	agttgcaatgagctcatgg	3803	3822	SEQ ID NO: 1203	ccalcagtcagataaact	7989	8008	15
SEQ ID NO: 201	tttgcaagaccacctcaat	3860	3879	SEQ ID NO: 1204	attgacctgtccattcaaa	13671	13690	15
SEQ ID NO: 202	gaaggagttcaacctccag	3884	3903	SEQ ID NO: 1205	ctggaattgtcattcttc	11728	11747	15

				NO:					
SEQ ID NO: 203	actccacatccagaaaa	3919	3938	SEQ ID NO: 1206	ttttaacaaaagtgaagt	6821	6840	15	
SEQ ID NO: 204	ctctcttaaaaagcgatg	3939	3958	SEQ ID NO: 1207	catcactgccaaaggagag	8486	8505	15	
SEQ ID NO: 205	aaaagcgatggcgggltca	3948	3967	SEQ ID NO: 1208	tgactcactcattgattt	12680	12699	15	
SEQ ID NO: 206	ttccttgcctttgtgtg	4003	4022	SEQ ID NO: 1209	ccacaaacaatgaagggaa	9256	9275	15	
SEQ ID NO: 207	caagctgtgggattccat	4079	4098	SEQ ID NO: 1210	atgggaaaaaacaggcttg	9566	9585	15	
SEQ ID NO: 208	aagtcctcattttaccat	4117	4136	SEQ ID NO: 1211	atgggaagataaagaactt	4834	4853	15	
SEQ ID NO: 209	tgctctctcgggtgtct	4159	4178	SEQ ID NO: 1212	agaaaaacaacacaggca	9643	9662	15	
SEQ ID NO: 210	accagcacagaccatttca	4242	4261	SEQ ID NO: 1213	tgaagtgtagtctcgtgt	5089	5108	15	
SEQ ID NO: 211	ccagcacagaccatttcag	4243	4262	SEQ ID NO: 1214	ctgaatacaatgctctgg	5511	5530	15	
SEQ ID NO: 212	actatcatgtgatgggtct	4367	4386	SEQ ID NO: 1215	agacactgattttatagt	7948	7967	15	
SEQ ID NO: 213	accagatgtctgtctca	4496	4515	SEQ ID NO: 1216	tgaaggctgactctgtgtt	4282	4301	15	
SEQ ID NO: 214	ccagatgtctgtctcag	4497	4516	SEQ ID NO: 1217	ctgagcaacaattgtgtg	10311	10330	15	
SEQ ID NO: 215	tttggactccaaaagaaa	4520	4539	SEQ ID NO: 1218	tttctcctgattacaaa	5933	5952	15	
SEQ ID NO: 216	tcaagaagaagcaagtga	4552	4571	SEQ ID NO: 1219	tcaaggataacgtgttga	12610	12629	15	
SEQ ID NO: 217	atgagaactacgagctgac	4798	4817	SEQ ID NO: 1220	gtcagatattgtgtcat	10187	10206	15	
SEQ ID NO: 218	ttaaatctgacaccaatg	4818	4837	SEQ ID NO: 1221	catctatgaagatgttaa	7342	7361	15	
SEQ ID NO: 219	gaagataagaactttgoc	4838	4857	SEQ ID NO: 1222	ggcaaatltgaaggacttc	11994	12013	15	
SEQ ID NO: 220	aagataagaactttgoca	4839	4858	SEQ ID NO: 1223	tggcaaatltgaaggactt	11993	12012	15	
SEQ ID NO: 221	ttcttcagctgctttctg	4941	4960	SEQ ID NO: 1224	cagaatccagatacaagaa	6884	6903	15	
SEQ ID NO: 222	ctggactcataaattocca	4957	4976	SEQ ID NO: 1225	tgggtcttccagagccag	11033	11052	15	
SEQ ID NO: 223	aaattaatagtggtctca	5014	5033	SEQ ID NO: 1226	tgagaagccccaagaattt	6248	6267	15	
SEQ ID NO: 224	agtgcaacgaccaactga	5073	5092	SEQ ID NO: 1227	tcaaatctcggatcacct	9848	9867	15	
SEQ ID NO: 225	ctgggaagtgtattcagg	5238	5257	SEQ ID NO: 1228	cctgaacctcataaccag	8310	8329	15	
SEQ ID NO: 226	gcaaaaacattttcaactt	5278	5297	SEQ ID NO: 1229	aagtaaaagaaaattttgc	10744	10763	15	
SEQ ID NO: 227	aaaaacattttcaacttca	5280	5299	SEQ ID NO: 1230	tgaagtaaaagaaaatttt	10742	10761	15	
SEQ ID NO: 228	tcagtcaagaaggacttaa	5302	5321	SEQ ID NO: 1231	ttaaggacttcattctga	13363	13382	15	
SEQ ID NO: 229	tcaaatgacatgatgggt	5325	5344	SEQ ID NO: 1232	agcccatcaatcatctga	6205	6224	15	
SEQ ID NO: 230	cacacaacagctgaaca	5367	5386	SEQ ID NO: 1233	tgttcaactgcctttgtg	11219	11238	15	

SEQ ID NO:	231	tcctcaaaacttgacaaca	5409	5428	SEQ ID NO:	1234	tgtttctatttccaaga	12835	12854	15
SEQ ID NO:	232	caagtttataagcaaact	5441	5460	SEQ ID NO:	1235	agttatttgcataacttg	14043	14062	15
SEQ ID NO:	233	tggttaactctttaaacag	5488	5507	SEQ ID NO:	1236	ctgttttagaggaaacca	7512	7531	15
SEQ ID NO:	234	aacagtgacctgaataaca	5502	5521	SEQ ID NO:	1237	tgtatagcaaatctcgtt	5890	5909	15
SEQ ID NO:	235	gggaactacggctagaac	5544	5563	SEQ ID NO:	1238	gttccttcacatgttccc	10933	10952	15
SEQ ID NO:	236	aacacatctatgccatctc	5620	5639	SEQ ID NO:	1239	gagacagcatctcgtgtt	11204	11223	15
SEQ ID NO:	237	tcagcaagctataaagcag	5652	5671	SEQ ID NO:	1240	ctgctaagaaccttactga	7780	7799	15
SEQ ID NO:	238	gcagacactgttgtaagg	5687	5686	SEQ ID NO:	1241	cccttcaagcactgactgc	11746	11765	15
SEQ ID NO:	239	tcggggagaacatactgg	5866	5885	SEQ ID NO:	1242	ccagggtttcacaccaga	8038	8057	15
SEQ ID NO:	240	ttctctcatgtattacaaag	5934	5953	SEQ ID NO:	1243	cttttccaccaaggagaa	10838	10857	15
SEQ ID NO:	241	ctgagcagacaggcacctg	6034	6053	SEQ ID NO:	1244	caggaggcttaagttcag	7599	7618	15
SEQ ID NO:	242	caatttaacaacaatgaat	6066	6085	SEQ ID NO:	1245	attccttctttacaattg	8082	8101	15
SEQ ID NO:	243	tggacgaactctggctgac	6140	6159	SEQ ID NO:	1246	glcagcccagttcttcca	10924	10943	15
SEQ ID NO:	244	cttttactcagtgagccca	6192	6211	SEQ ID NO:	1247	tgggctaaacgtatgaaag	7827	7846	15
SEQ ID NO:	245	tcaattgatgctttagagat	6217	6236	SEQ ID NO:	1248	atcttcataagttcaatga	13174	13193	15
SEQ ID NO:	246	aaaaccaagatgttcactc	6295	6314	SEQ ID NO:	1249	gagtgaaatcgtgttttt	8630	8649	15
SEQ ID NO:	247	aggaatcgacaaccatta	6357	6376	SEQ ID NO:	1250	taatgattttcaagttcct	8294	8313	15
SEQ ID NO:	248	tagttgtactggaaaacgt	6376	6395	SEQ ID NO:	1251	acgttagcctctaagacta	11928	11947	15
SEQ ID NO:	249	ggaaaacgtacagagaaag	6386	6405	SEQ ID NO:	1252	cttttacaattcattttcc	13014	13033	15
SEQ ID NO:	250	gaaaacgtacagagaaag	6387	6406	SEQ ID NO:	1253	gctttctctccacatttc	10052	10071	15
SEQ ID NO:	251	aaagctgaagcacatcaat	6401	6420	SEQ ID NO:	1254	attgatgttagatgcttt	6984	7003	15
SEQ ID NO:	252	aagctgaagcacatcaata	6402	6421	SEQ ID NO:	1255	tattgatgttagatgcttt	6983	7002	15
SEQ ID NO:	253	tgaagcacatcaataatga	6406	6425	SEQ ID NO:	1256	tcaaccttaagtatttca	8287	8306	15
SEQ ID NO:	254	atcaatattgcaatttg	6414	6433	SEQ ID NO:	1257	caaagccatcactgatgat	1660	1679	15
SEQ ID NO:	255	taagattatctgaattca	6476	6495	SEQ ID NO:	1258	tgaatcattgaaaaatta	6719	6738	15
SEQ ID NO:	256	gattatctgaattcattca	6480	6499	SEQ ID NO:	1259	tgaagtagctgagaaaaac	7094	7113	15
SEQ ID NO:	257	aattggagagacaagttt	6498	6517	SEQ ID NO:	1260	aaacattcctttaacaatt	9488	9507	15
SEQ ID NO:	258	aaaatagctattgctaata	6693	6712	SEQ ID NO:	1261	tattgaaatattgatttt	6806	6825	15

255

				NO:									
SEQ ID NO: 287	gggctgccccgattctc	8902	8921	SEQ ID NO:	1290	gagaacattatggaggccc	9432	9451	15				
SEQ ID NO: 288	ttctcagatgagggaacac	8916	8935	SEQ ID NO:	1291	gtgtcttcaagctgagaa	12408	12427	15				
SEQ ID NO: 289	gatgagggaaacatgaat	8922	8941	SEQ ID NO:	1292	attccagcttccccacatc	8330	8349	15				
SEQ ID NO: 290	ctttggactgtccaataag	8978	8997	SEQ ID NO:	1293	cttatgggatttctaag	11159	11178	15				
SEQ ID NO: 291	gcatacacaacaatgaag	9252	9271	SEQ ID NO:	1294	cttactgtctgattgatgc	10219	10238	15				
SEQ ID NO: 292	cacaacaatgaagggaat	9257	9276	SEQ ID NO:	1295	attccctgaagttgatgtg	11480	11499	15				
SEQ ID NO: 293	ccaaaatttctctgtgga	9407	9426	SEQ ID NO:	1296	ttcatcacaacattcttg	9663	9682	15				
SEQ ID NO: 294	caaaatttctctgtgga	9408	9427	SEQ ID NO:	1297	ttcatcacaacattcttg	9662	9681	15				
SEQ ID NO: 295	ctgtctggaacaacagaga	9417	9436	SEQ ID NO:	1298	ttcaagagttacagcaga	13221	13240	15				
SEQ ID NO: 296	ctgtctggaacaacagaga	9418	9437	SEQ ID NO:	1299	ttcaagagttacagcag	13220	13239	15				
SEQ ID NO: 297	agaacattatggaggccca	9433	9452	SEQ ID NO:	1300	tgggctgccccagattct	8901	8920	15				
SEQ ID NO: 298	agaagcaaatctggatttc	9467	9486	SEQ ID NO:	1301	gaaatctcaatttatct	13813	13832	15				
SEQ ID NO: 299	tttctctatgggaaaaa	9557	9576	SEQ ID NO:	1302	ttttgcaagttaaagaaa	14013	14032	15				
SEQ ID NO: 300	tcagagcatcaaatcttt	9704	9723	SEQ ID NO:	1303	aaagaaaatcaggatctga	14025	14044	15				
SEQ ID NO: 301	cagaacaatgcattagat	9743	9762	SEQ ID NO:	1304	atctatgcatctctctg	5625	5644	15				
SEQ ID NO: 302	tacacattaatctgccat	9993	10012	SEQ ID NO:	1305	atggagctttattgtgta	14081	14100	15				
SEQ ID NO: 303	agtcagatatgtgtcica	10186	10205	SEQ ID NO:	1306	tgaagactacgagctgact	4799	4818	15				
SEQ ID NO: 304	ggagggtatcataacagt	10328	10347	SEQ ID NO:	1307	actgttggtcaaacctccc	2726	2745	15				
SEQ ID NO: 305	caaaagocgaaattccat	10396	10415	SEQ ID NO:	1308	attgaagtacactctttg	8358	8377	15				
SEQ ID NO: 306	aaaagccgaattccaatt	10397	10416	SEQ ID NO:	1309	aattgaagtacactcttt	8357	8376	15				
SEQ ID NO: 307	ttcaagcaagaacttaag	10428	10447	SEQ ID NO:	1310	cattatggocctctgtgaa	13250	13269	15				
SEQ ID NO: 308	cctcttactttccattga	10570	10589	SEQ ID NO:	1311	tcaaaagaagccaagagg	12939	12958	15				
SEQ ID NO: 309	tgaggccaacacttaactg	10655	10674	SEQ ID NO:	1312	caagcatctgattgactca	12668	12687	15				
SEQ ID NO: 310	cacttactgaattccaag	10664	10683	SEQ ID NO:	1313	cttgaacacaaagtcagtg	6000	6019	15				
SEQ ID NO: 311	gaagtaaaagaaaatttg	10743	10762	SEQ ID NO:	1314	caaaaacattttcaacttc	5279	5298	15				
SEQ ID NO: 312	cctggaactctctccatgg	10874	10893	SEQ ID NO:	1315	ccattacagatcttcagg	11364	11383	15				
SEQ ID NO: 313	agctggatgttaaccacag	11176	11195	SEQ ID NO:	1316	ctggattccacatgcagct	11847	11866	15				
SEQ ID NO: 314	aaaattccctgaagttgat	11477	11496	SEQ ID NO:	1317	atcatatcogltgaatttt	6757	6776	15				

WU 2004/09/15/15				FC 17/03/2004/01/15			
SEQ ID NO:	315		NO:				
SEQ ID NO: 315	cagatggcattgctgcttt	11605	11624	SEQ ID NO: 1318	aaagctgagaagaatctg	12416	12435
SEQ ID NO: 316	agaatggcattgctgcttt	11606	11625	SEQ ID NO: 1319	caaagctgagaagaatct	12415	12434
SEQ ID NO: 317	tggtgaacagtcctggat	11834	11853	SEQ ID NO: 1320	atccaagatgagatcaaca	13095	13114
SEQ ID NO: 318	catattcaaaacagtggt	12221	12240	SEQ ID NO: 1321	caactctctgallactatg	13623	13642
SEQ ID NO: 319	aaagatttaicaaaagaag	12930	12949	SEQ ID NO: 1322	ctcaatttattctcttt	13818	13837
SEQ ID NO: 320	attttccaactaatagaag	13026	13045	SEQ ID NO: 1323	cttcaagacttaaaaaat	8006	8025
SEQ ID NO: 321	aattatataccaagatgaga	13089	13108	SEQ ID NO: 1324	tctcttctccatggaat	10471	10490
SEQ ID NO: 322	ttcaggaagcttctcaaga	13210	13229	SEQ ID NO: 1325	tcttcalaagttcaatgaa	13175	13194
SEQ ID NO: 323	ttgagcaattttcgcacag	13429	13448	SEQ ID NO: 1326	ctgttgaagatttatcaa	12924	12943
SEQ ID NO: 324	ctgatatcatcacggaggt	13704	13723	SEQ ID NO: 1327	actcaatggtgaaattcag	7457	7476
SEQ ID NO: 325	acatcacggagttactgaa	13711	13730	SEQ ID NO: 1328	ttcagaagctaagcaatgt	7231	7250
SEQ ID NO: 326	actgcctatattgataaaa	13874	13893	SEQ ID NO: 1329	ttttggcaagctatcacgt	8372	8391
SEQ ID NO: 327	aggatggcatttttgcac	14003	14022	SEQ ID NO: 1330	ttgcaagcaagctcttct	3005	3024
SEQ ID NO: 328	ttttttcgaagttaaagaa	14012	14031	SEQ ID NO: 1331	ttctctctatgggaaaaaa	9558	9577
SEQ ID NO: 329	tccagaactcaagicttca	1619	1638	SEQ ID NO: 1332	tgaatgcgtgttttttga	8633	8652
SEQ ID NO: 330	agtttagtgaagaagtctt	1948	1967	SEQ ID NO: 1333	agaatctgtaccaggaaat	12556	12575
SEQ ID NO: 331	attttacgctctgacaagt	5427	5446	SEQ ID NO: 1334	acttcagagaataacaaat	11401	11420
SEQ ID NO: 332	gattatctgaattcatcta	6480	6499	SEQ ID NO: 1335	tgaaccaatgcacaaaatc	7421	7440
SEQ ID NO: 333	gtgcctctctcggttgctg	18	37	SEQ ID NO: 1336	cagctgagcagacaggcac	6031	6050
SEQ ID NO: 334	attcaagcacctcoggag	245	264	SEQ ID NO: 1337	cttcataagttcaatgaat	13176	13195
SEQ ID NO: 335	gactgctgattcaagaagt	308	327	SEQ ID NO: 1338	acttccaactctcaagtc	13407	13426
SEQ ID NO: 336	ttgtcgcagcatgtccag	475	494	SEQ ID NO: 1339	cgggcagctgtalagcaa	5881	5900
SEQ ID NO: 337	agaaagatgaacctactta	547	566	SEQ ID NO: 1340	taagiatgatttcaattct	10490	10509
SEQ ID NO: 338	tgaagactctcagggaact	1087	1106	SEQ ID NO: 1341	agttcaatgaattatttca	13183	13202
SEQ ID NO: 339	atctctctgccacagctg	1202	1221	SEQ ID NO: 1342	cagcccagccatttgagat	9229	9248
SEQ ID NO: 340	tctctctgccacagctga	1203	1222	SEQ ID NO: 1343	tcagccagcagcttgaga	9228	9247
SEQ ID NO: 341	tgagggtccagcccatc	1223	1242	SEQ ID NO: 1344	gatgggaaagccgccctca	5208	5227
SEQ ID NO: 342	cogaactcaagttctcaa	1620	1639	SEQ ID NO: 1345	ttgaagcagaaacctctg	5907	5926

258

SEQ ID NO:	371	ctgagagttccagtgagtg	283	302	SEQ ID NO:	1374	actcacctggacattcag	3383	3402	14
SEQ ID NO:	372	tccagtgagtgctcctggga	291	310	SEQ ID NO:	1375	tccggagccaaggctgga	2675	2694	14
SEQ ID NO:	373	aggttgagctggaggttcc	346	365	SEQ ID NO:	1376	ggaacccctcctcacct	4728	4747	14
SEQ ID NO:	374	tgagctggaggttcccag	350	369	SEQ ID NO:	1377	ctggggagcatgatgctca	9163	9182	14
SEQ ID NO:	375	tctgcagcttcactctgaa	370	389	SEQ ID NO:	1378	ttcaataataatcggcaga	3261	3280	14
SEQ ID NO:	376	gccagtgacccctgaaaga	394	413	SEQ ID NO:	1379	tcttcggtctgtaatggc	5794	5813	14
SEQ ID NO:	377	ctctgaggagtttgctga	464	483	SEQ ID NO:	1380	tgcaagaataatttgagag	6340	6359	14
SEQ ID NO:	378	aggtatgagctcaagctgg	492	511	SEQ ID NO:	1381	ccagtttcggggaaacct	12716	12735	14
SEQ ID NO:	379	tctttaccggagaaaga	535	554	SEQ ID NO:	1382	tcttttgggaagcaagga	2219	2238	14
SEQ ID NO:	380	catcaagaggggcatcatt	575	594	SEQ ID NO:	1383	aatggcaagttcctgatg	2277	2296	14
SEQ ID NO:	381	tcttggttccccagagac	601	620	SEQ ID NO:	1384	gtctcgaactcagaagga	13988	14007	14
SEQ ID NO:	382	aagaagccaagcaagtggt	622	641	SEQ ID NO:	1385	aacaataaataggagtctt	14072	14091	14
SEQ ID NO:	383	aagcaagtggtttctgg	630	649	SEQ ID NO:	1386	ccagagccaggtcagagctt	11042	11061	14
SEQ ID NO:	384	tctgatacctgtatgga	644	663	SEQ ID NO:	1387	tccatgtccatttacaga	11358	11375	14
SEQ ID NO:	385	ccactcactttaccgtcaa	670	689	SEQ ID NO:	1388	ttgattttaacaagaagtg	6817	6836	14
SEQ ID NO:	386	aggaagggcaatgtgcaa	693	712	SEQ ID NO:	1389	ttgcaagcaagctttcct	3005	3024	14
SEQ ID NO:	387	gcaatgtggcaacagaaat	700	719	SEQ ID NO:	1390	atttccatcccctgttc	3480	3499	14
SEQ ID NO:	388	caatgtggcaacagaaata	701	720	SEQ ID NO:	1391	tattcttctttcaaatg	13826	13845	14
SEQ ID NO:	389	tggaacagaaatatccac	706	725	SEQ ID NO:	1392	gtggcttccatattgcca	1887	1906	14
SEQ ID NO:	390	agagacctgggccaagtgtg	729	748	SEQ ID NO:	1393	cacattacatttgtctct	2930	2949	14
SEQ ID NO:	391	tgtagctgtcaagccca	744	763	SEQ ID NO:	1394	tgggaaagccgcctcaca	5210	5229	14
SEQ ID NO:	392	gtgatgcttcaagcccat	745	764	SEQ ID NO:	1395	atgggaagccgcctcac	5209	5228	14
SEQ ID NO:	393	cagoccatgtctctcatc	776	795	SEQ ID NO:	1396	gatgctgaacagtgaagctg	8144	8163	14
SEQ ID NO:	394	gctctcatcaaggcatga	786	805	SEQ ID NO:	1397	tcataacagtactgtgagc	10337	10356	14
SEQ ID NO:	395	cctgtcaactctgatcag	811	830	SEQ ID NO:	1398	ctgagtggtttatcaag	12445	12464	14
SEQ ID NO:	396	cttgcaactctgatcagc	812	831	SEQ ID NO:	1399	gctgagtggtttatcaag	12444	12463	14
SEQ ID NO:	397	agccatctgcaaggagcaa	884	903	SEQ ID NO:	1400	ttgcaatgagctcatggct	3805	3824	14
SEQ ID NO:	398	gccatctgcaaggagcaac	885	904	SEQ ID NO:	1401	gttgcaatgagctcatggc	3804	3823	14

SEQ ID NO:	399	ctctcgtcttctctac	908	927	NO:	1402	gtaggataaatggagaag	9453	9472	14
SEQ ID NO:	400	ctttctctacaagaataa	916	935	SEQ ID NO:	1403	ttattgctgaatcaaaag	13648	13667	14
SEQ ID NO:	401	gatcaacagcgcttcttt	989	1008	SEQ ID NO:	1404	aaagccatcactgatgac	1661	1680	14
SEQ ID NO:	402	atcaacagcgcttctttg	990	1009	SEQ ID NO:	1405	caaagccatcactgatgat	1660	1679	14
SEQ ID NO:	403	acagcgcttctttgttga	994	1013	SEQ ID NO:	1406	tcacaaatcctttggctgt	9667	9686	14
SEQ ID NO:	404	aagatggcgctcgactttg	1023	1042	SEQ ID NO:	1407	caaaatagaagggaactctt	2069	2088	14
SEQ ID NO:	405	tgttttgaagactctccag	1082	1101	SEQ ID NO:	1408	ctgglaactactttaaaca	5487	5506	14
SEQ ID NO:	406	ttaagactctccaggaaac	1086	1105	SEQ ID NO:	1409	gttcaatgaattatttcaa	13184	13203	14
SEQ ID NO:	407	aactgaaaaaactaacat	1102	1121	SEQ ID NO:	1410	atggcatlitttgcaagt	14006	14025	14
SEQ ID NO:	408	ctgaaaaaactaacactct	1104	1123	SEQ ID NO:	1411	agattgatgggcagttcag	4564	4583	14
SEQ ID NO:	409	aaaactaaccatctctgag	1109	1128	SEQ ID NO:	1412	ctcaagaatgactttttt	2570	2589	14
SEQ ID NO:	410	tgaacaaatatccagaga	1124	1143	SEQ ID NO:	1413	ttccagataaaaaactca	12201	12220	14
SEQ ID NO:	411	caataagctggttactgag	1154	1173	SEQ ID NO:	1414	ctcagctcaaatgaatttg	12265	12284	14
SEQ ID NO:	412	tactgagctgagagccctc	1166	1185	SEQ ID NO:	1415	gagggtgctcataacagta	10329	10348	14
SEQ ID NO:	413	gctcagtgatgaagcagt	1180	1199	SEQ ID NO:	1416	actgttgactcaggaaaggc	12572	12591	14
SEQ ID NO:	414	agtcacatctcttgcoca	1196	1215	SEQ ID NO:	1417	tggccacatagctggact	8858	8877	14
SEQ ID NO:	415	atctcttggccacagctg	1202	1221	SEQ ID NO:	1418	cagctgacctcatcgagat	2161	2180	14
SEQ ID NO:	416	ttctcttgccacagctga	1203	1222	SEQ ID NO:	1419	tcagctgacctcatcgaga	2160	2179	14
SEQ ID NO:	417	tggcacagctgattgaggt	1210	1229	SEQ ID NO:	1420	aactgcacaaagctggca	13955	13974	14
SEQ ID NO:	418	gccacagctgattgagtg	1211	1230	SEQ ID NO:	1421	caccaaaacccaatggc	11240	11259	14
SEQ ID NO:	419	tcactttacaagccttgg	1240	1259	SEQ ID NO:	1422	accagatgctgaacagtga	8140	8159	14
SEQ ID NO:	420	ccctctgatagatgttgt	1324	1343	SEQ ID NO:	1423	accacttacagctagaggg	10816	10835	14
SEQ ID NO:	421	gtcacctacctggtggccc	1341	1360	SEQ ID NO:	1424	gggcgacctaatgtgtgac	3431	3450	14
SEQ ID NO:	422	cttgtatgcgctgagcca	1432	1451	SEQ ID NO:	1425	tggctgttaacctaaaagg	5578	5597	14
SEQ ID NO:	423	gacaaacctacagggacc	1472	1491	SEQ ID NO:	1426	ggctcttatgattatgtc	12347	12366	14
SEQ ID NO:	424	tgtcaattacctgatggaa	1508	1527	SEQ ID NO:	1427	ttccaaaagcagtcagca	9930	9949	14
SEQ ID NO:	425	tgactgcactggggatgaa	1538	1557	SEQ ID NO:	1428	ttcaggctcatgcaagtca	10909	10928	14
SEQ ID NO:	426	actgcactggggatgaaga	1540	1559	SEQ ID NO:	1429	tcttgaacacaaagtcagt	5999	6018	14

SEQ ID NO: 427	atgaagattacaactattt	1552	1571	NO:	SEQ ID NO: 1430	aaatgaagtaaagatcat	8110	8129	14
SEQ ID NO: 428	accatggagcagtaactc	1602	1621	SEQ ID NO:	SEQ ID NO: 1431	gagtaaacacaaacttgt	9016	9035	14
SEQ ID NO: 429	gcagttacccagaaactc	1610	1629	SEQ ID NO:	SEQ ID NO: 1432	gagttactgaaaaagctgc	13719	13738	14
SEQ ID NO: 430	cagaactcaagttctcaat	1621	1640	SEQ ID NO:	SEQ ID NO: 1433	attgatatccaagatctg	1925	1944	14
SEQ ID NO: 431	caggctctgcggaatagg	1695	1714	SEQ ID NO:	SEQ ID NO: 1434	ccatgacctccagctctg	2477	2496	14
SEQ ID NO: 432	ccaggagggtctctcag	1730	1749	SEQ ID NO:	SEQ ID NO: 1435	ctgaaatacaatgclctgg	5511	5530	14
SEQ ID NO: 433	gggtctctcagacttc	1736	1755	SEQ ID NO:	SEQ ID NO: 1436	gaaaaacttggaaacaacc	4431	4450	14
SEQ ID NO: 434	tttcttgatgatgttct	1751	1770	SEQ ID NO:	SEQ ID NO: 1437	agaatccagatacaagaaa	6885	6904	14
SEQ ID NO: 435	ggagataagcagtgctg	1773	1792	SEQ ID NO:	SEQ ID NO: 1438	cagcatgcctagtctcc	9944	9963	14
SEQ ID NO: 436	gctgcctactatgttga	1788	1807	SEQ ID NO:	SEQ ID NO: 1439	tcaatatcaaaagcccagc	12037	12056	14
SEQ ID NO: 437	acttggtgctccat	1882	1901	SEQ ID NO:	SEQ ID NO: 1440	atatctggaacctgaagt	10729	10748	14
SEQ ID NO: 438	gccaatatctgaactcag	1902	1921	SEQ ID NO:	SEQ ID NO: 1441	ctgaactcagaaggatggc	13992	14011	14
SEQ ID NO: 439	aatatcttgaactcagaag	1905	1924	SEQ ID NO:	SEQ ID NO: 1442	cttcattctgaatatatt	13370	13389	14
SEQ ID NO: 440	ctcagaagaattggatgc	1916	1935	SEQ ID NO:	SEQ ID NO: 1443	gataaaagattacttgag	7265	7284	14
SEQ ID NO: 441	aagaattggatccaaga	1921	1940	SEQ ID NO:	SEQ ID NO: 1444	tcttcaatttatctctct	13817	13836	14
SEQ ID NO: 442	agaattggatccaagat	1922	1941	SEQ ID NO:	SEQ ID NO: 1445	atctcaatttatctctct	13816	13835	14
SEQ ID NO: 443	tggatatccaagctcgaa	1927	1946	SEQ ID NO:	SEQ ID NO: 1446	ttcacataccaagaattcca	8317	8336	14
SEQ ID NO: 444	atatccaagatcgaaaaa	1930	1949	SEQ ID NO:	SEQ ID NO: 1447	tttttaaccagtcagatat	10177	10196	14
SEQ ID NO: 445	tatccaagatcgaaaaag	1931	1950	SEQ ID NO:	SEQ ID NO: 1448	cttttaaccagtcagata	10176	10195	14
SEQ ID NO: 446	caagatcgaaaaagttag	1935	1954	SEQ ID NO:	SEQ ID NO: 1449	ctaaattccatgggtctg	4965	4984	14
SEQ ID NO: 447	aagatcgaaaaagttag	1936	1955	SEQ ID NO:	SEQ ID NO: 1450	actaaattccatgggtctt	4964	4983	14
SEQ ID NO: 448	tgaaaaagttagtgaaga	1942	1961	SEQ ID NO:	SEQ ID NO: 1451	tctttctgggaatatca	10622	10641	14
SEQ ID NO: 449	tccaactgcatggacttc	1982	2001	SEQ ID NO:	SEQ ID NO: 1452	gaagcacatatgaactgga	13937	13956	14
SEQ ID NO: 450	tcagaaaattctctcgaa	1999	2018	SEQ ID NO:	SEQ ID NO: 1453	ttcctttaacaattcctga	9493	9512	14
SEQ ID NO: 451	ttcatcacttgaccagc	2044	2063	SEQ ID NO:	SEQ ID NO: 1454	gctgacataggaatggaa	8433	8452	14
SEQ ID NO: 452	ccagcctcagccaaaata	2057	2076	SEQ ID NO:	SEQ ID NO: 1455	tattctatccaagattggg	7812	7831	14
SEQ ID NO: 453	agcctcagccaaaatagaa	2060	2079	SEQ ID NO:	SEQ ID NO: 1456	ttctatccaagattgggt	7814	7833	14
SEQ ID NO: 454	atcttatattgatccaaa	2083	2102	SEQ ID NO:	SEQ ID NO: 1457	ttgaaaaacaagcagat	11813	11832	14

SEQ ID NO: 455	tcttatattgatccaaat	2084	2103	SEQ ID NO: 455	1458	atttttgcagattaaaga	14011	14030	14
SEQ ID NO: 456	cttcttaagaaagcatgc	2109	2128	SEQ ID NO: 456	1459	gcatggcattatgatgaag	3806	3625	14
SEQ ID NO: 457	ctaagaaagcatgctgaa	2113	2132	SEQ ID NO: 457	1460	ttcagggtgtgagtttag	5686	5705	14
SEQ ID NO: 458	taagaaagcatgctgaaa	2114	2133	SEQ ID NO: 458	1461	tttctaaccattccttta	9482	9501	14
SEQ ID NO: 459	gagattggcttggaggaa	2175	2194	SEQ ID NO: 459	1462	ttccctccattagtctc	11701	11720	14
SEQ ID NO: 460	ctttgaccaacattggaa	2198	2217	SEQ ID NO: 460	1463	ttcaattgaccaagaaaag	11060	11079	14
SEQ ID NO: 461	cagacagtgtcaacaagc	2245	2264	SEQ ID NO: 461	1464	gcttactggaagcaactctg	6134	6153	14
SEQ ID NO: 462	cagtgctcaacaagctttg	2249	2268	SEQ ID NO: 462	1465	caaattctcgatacactg	9849	9868	14
SEQ ID NO: 463	agtgcaacaagctttgt	2250	2269	SEQ ID NO: 463	1466	acaagaaacgtctacact	4351	4370	14
SEQ ID NO: 464	ctgatgggtctctaaggt	2290	2309	SEQ ID NO: 464	1467	acctcggaacaatcctcag	3325	3344	14
SEQ ID NO: 465	tgatgggtctctaaggtc	2291	2310	SEQ ID NO: 465	1468	gacctcgcaacagatgca	8823	8842	14
SEQ ID NO: 466	aaacatgagcaggatag	2343	2362	SEQ ID NO: 466	1469	ccatgatctacattgttt	6788	6807	14
SEQ ID NO: 467	gaagctgattaaagattg	2387	2406	SEQ ID NO: 467	1470	caaaaaattttcaacttc	5279	5298	14
SEQ ID NO: 468	aaagattgaaatccaag	2397	2416	SEQ ID NO: 468	1471	ctttaagtctcagcatctt	7806	7825	14
SEQ ID NO: 469	gatgggtgcccgcactctg	2510	2529	SEQ ID NO: 469	1472	cagattgaggattccatc	7975	7994	14
SEQ ID NO: 470	gggatccccagatgattg	2532	2551	SEQ ID NO: 470	1473	caatcacaagctgattccc	9075	9094	14
SEQ ID NO: 471	ttttctcactacatcttc	2585	2604	SEQ ID NO: 471	1474	gaagtgctcagtgcaaaaa	10374	10393	14
SEQ ID NO: 472	tcttctcactacatctcatg	2588	2607	SEQ ID NO: 472	1475	catggcattatgatgaaga	3607	3626	14
SEQ ID NO: 473	tacatctcatcgagaaatg	2595	2614	SEQ ID NO: 473	1476	cattatggaggcccatgta	9437	9456	14
SEQ ID NO: 474	ttcatggagaatgcctttg	2601	2620	SEQ ID NO: 474	1477	caaaatcaactttaatgaa	6599	6618	14
SEQ ID NO: 475	tcatggagaatgcctttga	2602	2621	SEQ ID NO: 475	1478	tcaacacaacttcaatga	13108	13127	14
SEQ ID NO: 476	tttgaactccccactgag	2616	2635	SEQ ID NO: 476	1479	ctccccaggacctttcaaa	9834	9853	14
SEQ ID NO: 477	ttgaactccccactggagc	2617	2636	SEQ ID NO: 477	1480	gctccccaggacctttcaa	9833	9852	14
SEQ ID NO: 478	tgaactccccactggagct	2618	2637	SEQ ID NO: 478	1481	agctccccaggacctttca	9832	9851	14
SEQ ID NO: 479	cactggagctgattacag	2627	2646	SEQ ID NO: 479	1482	ctgttctgagtcocagtg	9336	9355	14
SEQ ID NO: 480	actggagctgattacagt	2628	2647	SEQ ID NO: 480	1483	actgttctgagtcocagtg	9335	9354	14
SEQ ID NO: 481	agttgcaaatatctcatc	2644	2663	SEQ ID NO: 481	1484	gatgatgccaaaatcaact	6591	6610	14
SEQ ID NO: 482	gttgcaaatatctcatct	2645	2664	SEQ ID NO: 482	1485	agatgatgccaaaatcaac	6590	6609	14

SEQ ID NO: 483	aaatactctcatctggagt	2650	2669	SEQ ID NO: 1486	actcagaaggatggcattt	13996	14015	14
SEQ ID NO: 484	taaaactggaagtagccaa	2695	2714	SEQ ID NO: 1487	ttggttacaggaggcttta	7592	7611	14
SEQ ID NO: 485	ggctgaactgggtggcaaaa	2720	2739	SEQ ID NO: 1488	tttctttcagccagacc	9220	9239	14
SEQ ID NO: 486	tgtggagtttgacaaat	2750	2769	SEQ ID NO: 1489	attttcaagcaaatgcaca	8530	8549	14
SEQ ID NO: 487	tttgacaatatgggcat	2758	2777	SEQ ID NO: 1490	atcgctctacctacacaa	9513	9532	14
SEQ ID NO: 488	atgaacaccaactcttcc	2811	2830	SEQ ID NO: 1491	ggaagctgaagttatcat	2869	2888	14
SEQ ID NO: 489	cttcacagctgggctctg	2825	2844	SEQ ID NO: 1492	cagagctactcctgggaag	5227	5246	14
SEQ ID NO: 490	gagtcgggtctggaggctc	2832	2851	SEQ ID NO: 1493	gagctactggagcaactc	6132	6151	14
SEQ ID NO: 491	cctaaaagctgggaagctg	2858	2877	SEQ ID NO: 1494	cagcctcccagcgtagg	12112	12131	14
SEQ ID NO: 492	agctgggaagctgaagtt	2884	2883	SEQ ID NO: 1495	aaactgttaattacagct	5455	5474	14
SEQ ID NO: 493	ccagattagagctggaact	3106	3125	SEQ ID NO: 1496	agttccggggaaacctgg	12718	12737	14
SEQ ID NO: 494	ggataccctgaagtttgta	3200	3219	SEQ ID NO: 1497	tacagtatctgaaatcc	8385	8404	14
SEQ ID NO: 495	ctgggctaccalgcatt	3244	3263	SEQ ID NO: 1498	aatgagctatggctcag	3809	3828	14
SEQ ID NO: 496	tgctcagtgaaagccaat	3289	3308	SEQ ID NO: 1499	attttgagaggaatcgaca	6349	6368	14
SEQ ID NO: 497	aattccggallttgatgtt	3305	3324	SEQ ID NO: 1500	aacacatgaatcacaaat	8930	8949	14
SEQ ID NO: 498	ttccgattttgatgtga	3307	3326	SEQ ID NO: 1501	tcaaaacgagctcaggaa	13199	13218	14
SEQ ID NO: 499	oggaacaatcctcagagtt	3329	3348	SEQ ID NO: 1502	aactgtlaacaactgctcg	4203	4222	14
SEQ ID NO: 500	tctcagagttatgatga	3337	3356	SEQ ID NO: 1503	tcatcaattggttacagga	7585	7604	14
SEQ ID NO: 501	ctcaccctggacattcaga	3384	3403	SEQ ID NO: 1504	ctcgagaacaatgctgag	12431	12450	14
SEQ ID NO: 502	cattcagaacaagaaaatt	3395	3414	SEQ ID NO: 1505	aattgactttgtagaattg	8096	8115	14
SEQ ID NO: 503	actgaggtcgccctcatgg	3414	3433	SEQ ID NO: 1506	ccatgcaagtcagccagat	10916	10935	14
SEQ ID NO: 504	ttatttcatacccggtt	3478	3497	SEQ ID NO: 1507	aaactgcctatattgataa	13872	13891	14
SEQ ID NO: 505	gtttgcaagcagaagccag	3493	3512	SEQ ID NO: 1508	ctggactctctcctcaaac	5400	5419	14
SEQ ID NO: 506	tttgcaagcagaagccaga	3494	3513	SEQ ID NO: 1509	ctcggtgtcgacagacaa	5264	5283	14
SEQ ID NO: 507	ttgcaagcagaagccagaa	3495	3514	SEQ ID NO: 1510	ttcggtgtcgacagacaa	5263	5282	14
SEQ ID NO: 508	ctgctctcctcaatggact	3546	3565	SEQ ID NO: 1511	agtcagattgatggcag	4559	4578	14
SEQ ID NO: 509	tgctacagcttatggctcc	3569	3588	SEQ ID NO: 1512	ggaggcttaagtgcaga	7601	7620	14
SEQ ID NO: 510	acagcttatggctccacag	3573	3592	SEQ ID NO: 1513	ctglatagcaaatcctgt	5889	5908	14

[illegible]

SEQ ID NO: 567	aaacattttcaactcaa	5281	5300	NO:		SEQ ID NO: 1570	ttgatgttagagctgttt	6985	7004	14
SEQ ID NO: 568	cttaagctctcaaatgaca	5316	5335	SEQ ID NO:		SEQ ID NO: 1571	tgctctacaacaagtaag	7247	7266	14
SEQ ID NO: 569	ttaagctctcaaatgacat	5317	5336	SEQ ID NO:		SEQ ID NO: 1572	atgtctacaacaagtaa	7246	7265	14
SEQ ID NO: 570	catgatgggtcatatgct	5333	5352	SEQ ID NO:		SEQ ID NO: 1573	agcatcttggtctcacatg	7616	7635	14
SEQ ID NO: 571	tggtgctcatgctgaat	5338	5357	SEQ ID NO:		SEQ ID NO: 1574	atttatcaaaagagccca	12934	12953	14
SEQ ID NO: 572	actggactctctcaaaa	5399	5418	SEQ ID NO:		SEQ ID NO: 1575	tttggcaagctatacagt	8372	8391	14
SEQ ID NO: 573	actctctcaaaactga	5404	5423	SEQ ID NO:		SEQ ID NO: 1576	tcaattggagagacaagt	6496	6515	14
SEQ ID NO: 574	ctgacaagttttataagca	5437	5456	SEQ ID NO:		SEQ ID NO: 1577	tgcttgtgagtattcag	9685	9704	14
SEQ ID NO: 575	aagttttataagcaaacg	5442	5461	SEQ ID NO:		SEQ ID NO: 1578	cagtcactgtagaaaactt	4421	4440	14
SEQ ID NO: 576	ctgttaattacagctaca	5458	5477	SEQ ID NO:		SEQ ID NO: 1579	tgtactggaaacgctacag	6380	6399	14
SEQ ID NO: 577	ttacagctacagccatt	5466	5485	SEQ ID NO:		SEQ ID NO: 1580	aatatgtatcaatttgtaa	6417	6436	14
SEQ ID NO: 578	ctggtaactactttaaac	5486	5505	SEQ ID NO:		SEQ ID NO: 1581	gtttgaaaaacaagcaga	11812	11831	14
SEQ ID NO: 579	tttaaacagtgaacctgaaa	5498	5517	SEQ ID NO:		SEQ ID NO: 1582	tttacttgaagaataaaa	7024	7043	14
SEQ ID NO: 580	ttaacagtgacctgaat	5499	5518	SEQ ID NO:		SEQ ID NO: 1583	atttcaagcaagaacttaa	10426	10445	14
SEQ ID NO: 581	cagtgacctgaatacaat	5504	5523	SEQ ID NO:		SEQ ID NO: 1584	attggcgtggagctactg	6123	6142	14
SEQ ID NO: 582	tgtggcgtgtaacctaaaa	5576	5595	SEQ ID NO:		SEQ ID NO: 1585	tttgcgtggagaaccaca	10757	10776	14
SEQ ID NO: 583	ttatcagcaagctataaag	5649	5668	SEQ ID NO:		SEQ ID NO: 1586	ctttgcactatgttctaaa	12756	12775	14
SEQ ID NO: 584	ggttcagggtgtggagttt	5684	5703	SEQ ID NO:		SEQ ID NO: 1587	aaacacctaaagagtaaac	9006	9025	14
SEQ ID NO: 585	attcagactcactgcattt	5767	5786	SEQ ID NO:		SEQ ID NO: 1588	aaatgctgacatagggaa	8429	8448	14
SEQ ID NO: 586	ttcagactcactgcatttc	5768	5787	SEQ ID NO:		SEQ ID NO: 1589	gaaatattatgaacttgaa	13304	13323	14
SEQ ID NO: 587	tacaaatggcaatgggaaa	5840	5859	SEQ ID NO:		SEQ ID NO: 1590	tttctaagactggagtga	11168	11187	14
SEQ ID NO: 588	gctgtatagcaaatctcg	5888	5907	SEQ ID NO:		SEQ ID NO: 1591	caggctcatgcaagtcagc	10911	10930	14
SEQ ID NO: 589	tgagcagacaggcacctgg	6035	6054	SEQ ID NO:		SEQ ID NO: 1592	ccagcttcccaacatctca	8333	8352	14
SEQ ID NO: 590	ggcaccttgaaactcaaga	6045	6064	SEQ ID NO:		SEQ ID NO: 1593	tcttggtttcaactgcc	11213	11232	14
SEQ ID NO: 591	tgaatacagccaggactgt	6080	6099	SEQ ID NO:		SEQ ID NO: 1594	caagtaagtgtcaggttca	9372	9391	14
SEQ ID NO: 592	gaatacagccaggactgtg	6081	6100	SEQ ID NO:		SEQ ID NO: 1595	ccaacactactgaattc	10660	10679	14
SEQ ID NO: 593	ctggacgaactctggctga	6139	6158	SEQ ID NO:		SEQ ID NO: 1596	tcagaaagctacctccag	7931	7950	14
SEQ ID NO: 594	ttttactcagtgaagccat	6193	6212	SEQ ID NO:		SEQ ID NO: 1597	atggactcttctggaaaa	8870	8889	14

267

SEQ ID NO: 623	tgaaggagactattcagaa	7219	7238	SEQ ID NO: 1626	ttcctaatttgatttca	8518	8537	14
SEQ ID NO: 624	gagactatcagaagctaa	7224	7243	SEQ ID NO: 1627	ttagccacagctctgtctc	10293	10312	14
SEQ ID NO: 625	aattagttgattattga	7285	7304	SEQ ID NO: 1628	tcaagaagcttaataaatt	7312	7331	14
SEQ ID NO: 626	gcttaalgaattatcttt	7319	7338	SEQ ID NO: 1629	aaaacagctcaggaagc	13201	13220	14
SEQ ID NO: 627	ttaacaaactcctgacat	7357	7376	SEQ ID NO: 1630	atgtctacacaaagtaa	7246	7265	14
SEQ ID NO: 628	aaattaaagtcatttgatt	7386	7405	SEQ ID NO: 1631	aatcctttgacaggcattt	9715	9734	14
SEQ ID NO: 629	gactcaatgggtaaatca	7456	7475	SEQ ID NO: 1632	tgaattcaatcacaaagtc	9068	9087	14
SEQ ID NO: 630	gaattcaggctctgggac	7467	7486	SEQ ID NO: 1633	gttctcaattttgatttc	8517	8536	14
SEQ ID NO: 631	actaccacaaaaagctgaa	7484	7503	SEQ ID NO: 1634	ttcaggaactattgctagt	10637	10656	14
SEQ ID NO: 632	ccaaaataaccttaactcat	7570	7589	SEQ ID NO: 1635	atgattccctgaacctgg	10942	10961	14
SEQ ID NO: 633	aaataacctaatacatcaa	7573	7592	SEQ ID NO: 1636	tgaagtaaaaaaaattt	10741	10760	14
SEQ ID NO: 634	tttaagtcagcatctttg	7607	7626	SEQ ID NO: 1637	caaacttggtattctttaa	9472	9491	14
SEQ ID NO: 635	cagggttatagcacacttg	7731	7750	SEQ ID NO: 1638	caagggttactgttctcg	7857	7876	14
SEQ ID NO: 636	gttactgttctgaaatc	7862	7881	SEQ ID NO: 1639	gattctcagatgagggaac	8914	8933	14
SEQ ID NO: 637	cactgttctgaaatcaag	7865	7884	SEQ ID NO: 1640	cttgaacacaaagtcagtg	6000	6019	14
SEQ ID NO: 638	actgttctgaaatcaaga	7866	7885	SEQ ID NO: 1641	tcttgaacacaaagtcagt	5999	6018	14
SEQ ID NO: 639	gcctgcctttgaagtcagt	7901	7920	SEQ ID NO: 1642	actgttgactcaggaaggc	12572	12591	14
SEQ ID NO: 640	taacagatttgaggattcc	7972	7991	SEQ ID NO: 1643	ggaagcttctcaagagtta	13214	13233	14
SEQ ID NO: 641	gttttccacaccagaattt	8042	8061	SEQ ID NO: 1644	aaatttctctgtcggaac	9410	9429	14
SEQ ID NO: 642	tcagaaccattgaccagat	8128	8147	SEQ ID NO: 1645	atctgcagaacaatgctga	12430	12449	14
SEQ ID NO: 643	tagcgagaatcacctgcc	8218	8237	SEQ ID NO: 1646	ggcagcttctgtgctgcta	12293	12312	14
SEQ ID NO: 644	ccitaatgatttcaagtt	8291	8310	SEQ ID NO: 1647	aactgttgactcaggaagg	12571	12590	14
SEQ ID NO: 645	acataccagaattccagct	8320	8339	SEQ ID NO: 1648	agctgcagcttctcatgt	10018	10037	14
SEQ ID NO: 646	aatgctgacatagggaatg	8430	8449	SEQ ID NO: 1649	cattatctctgccatcatt	9997	10016	14
SEQ ID NO: 647	atgctgacatagggaatgg	8431	8450	SEQ ID NO: 1650	ccatttgagatcacggcat	9237	9256	14
SEQ ID NO: 648	aaacacctcagcaaacgaa	8450	8469	SEQ ID NO: 1651	tttgttttccataagggt	9283	9302	14
SEQ ID NO: 649	agcaggatctgcagcttcc	8468	8487	SEQ ID NO: 1652	ggaagtggccctgaatgct	10964	10983	14
SEQ ID NO: 650	tgcacaactctcaaacctt	8543	8562	SEQ ID NO: 1653	agggaaagagaagattgca	13493	13512	14

				NO:					
SEQ ID NO: 651	aggagtcagtgagttctc	3584	3603	SEQ ID NO: 1654	gagaactactatcatcct	13780	13799	14	
SEQ ID NO: 652	tttttggaatgccattga	3644	3663	SEQ ID NO: 1655	tcaatgaattattcaaaa	13186	13205	14	
SEQ ID NO: 653	aatggagtgatgtcaaga	3721	3740	SEQ ID NO: 1656	tccttcagccagccatt	9223	9242	14	
SEQ ID NO: 654	gtcaagataacaatcagc	3733	3752	SEQ ID NO: 1657	gctgactttaaactcgac	4811	4830	14	
SEQ ID NO: 655	tcacaatgaacatccc	3779	3798	SEQ ID NO: 1658	gggattctcctaagctgga	11164	11183	14	
SEQ ID NO: 656	tgaacatcccaaacctgg	3787	3806	SEQ ID NO: 1659	ccagttccaggagctcaa	12595	12614	14	
SEQ ID NO: 657	acatcccaaacctggactt	3791	3810	SEQ ID NO: 1660	aagtcgattcccagctgt	9082	9101	14	
SEQ ID NO: 658	acttctctagtcaggctga	3806	3825	SEQ ID NO: 1661	tcagatggaaaaatgaagt	11002	11021	14	
SEQ ID NO: 659	tgaatcacaaattagtttc	3936	3955	SEQ ID NO: 1662	gaaagtcataatggltca	12809	12828	14	
SEQ ID NO: 660	agaaggacccctacttcc	3960	3979	SEQ ID NO: 1663	ggaagaagaggcagcttct	12284	12303	14	
SEQ ID NO: 661	ttggactgtccaataagat	3980	3999	SEQ ID NO: 1664	atcataatgcagtaccga	11626	11645	14	
SEQ ID NO: 662	actgtccaataagatcaat	3984	4003	SEQ ID NO: 1665	atgataaaaccatacagt	13883	13902	14	
SEQ ID NO: 663	ctgtccaataagatcaata	3985	4004	SEQ ID NO: 1666	tattgataaaaccatacag	13882	13901	14	
SEQ ID NO: 664	gtttatgaactgtggtccc	9033	9052	SEQ ID NO: 1667	gggaatctgatgaggaaac	12247	12266	14	
SEQ ID NO: 665	atgaatctgtgctccctcaa	9037	9056	SEQ ID NO: 1668	ttgattgtccaccatcat	11659	11678	14	
SEQ ID NO: 666	ctcaacttttctaacttg	9051	9070	SEQ ID NO: 1669	caagatcgagactttgag	11645	11664	14	
SEQ ID NO: 667	ctaaaggcatggcactgtt	9121	9140	SEQ ID NO: 1670	aacagaacaatgcattag	9741	9760	14	
SEQ ID NO: 668	aaggcatggcactgtttgg	9124	9143	SEQ ID NO: 1671	ccaagaagaaggcacacctt	11069	11088	14	
SEQ ID NO: 669	atccacaacaatgaagg	9254	9273	SEQ ID NO: 1672	ccctaacagatttgaggat	7969	7988	14	
SEQ ID NO: 670	ggaattgaaagtctgttt	9271	9290	SEQ ID NO: 1673	aaacaacacaggcattcc	9647	9666	14	
SEQ ID NO: 671	aataactatgcactgtttc	9324	9343	SEQ ID NO: 1674	gaaatactgtttctatt	12828	12847	14	
SEQ ID NO: 672	gaaacaacgagaacattat	9424	9443	SEQ ID NO: 1675	ataaactgcaagattttc	13600	13619	14	
SEQ ID NO: 673	ttcttgaaaacgacaaagc	9591	9610	SEQ ID NO: 1676	gcttccaatgaccaagaa	11057	11076	14	
SEQ ID NO: 674	ataagaaaaacaacacag	9640	9659	SEQ ID NO: 1677	ctgtgcttgtgagtttat	9682	9701	14	
SEQ ID NO: 675	aaaacaacacaggcattc	9646	9665	SEQ ID NO: 1678	gaattgaaagtctgttt	9272	9291	14	
SEQ ID NO: 676	gcattccatcacaaatcct	9659	9678	SEQ ID NO: 1679	aggaaatggccctgaatgc	10963	10982	14	
SEQ ID NO: 677	tttgaaaaaaacagaaca	9732	9751	SEQ ID NO: 1680	tggtgaagatttatcaaa	12925	12944	14	
SEQ ID NO: 678	caatgcattagattttgtc	9749	9768	SEQ ID NO: 1681	gacaagaaaaaggggattg	10271	10290	14	

SEQ ID NO:	679	caaagctgaaaaatctcag	9809	9828	SEQ ID NO:	1682	ctgagaactctcatcttg	11430	11449	14
SEQ ID NO:	680	ctcgatgatacactgttcag	9855	9874	SEQ ID NO:	1683	ctggactctctagtcagg	8802	8821	14
SEQ ID NO:	681	gttgaagtgtctccattca	9882	9901	SEQ ID NO:	1684	tgaatctgctccctcaac	9038	9057	14
SEQ ID NO:	682	tttctcatctaggttct	9956	9975	SEQ ID NO:	1685	agaatccagatacaagaaa	6885	6904	14
SEQ ID NO:	683	ttctcatctaggttctg	9957	9976	SEQ ID NO:	1686	cagaatccagatacaagaa	6884	6903	14
SEQ ID NO:	684	tcattagactgcagctcc	10011	10030	SEQ ID NO:	1687	ggacagtgaatattatga	13297	13316	14
SEQ ID NO:	685	tgtcgaaacttttaaccag	10169	10188	SEQ ID NO:	1688	ctggatgtaaccaccagca	11178	11197	14
SEQ ID NO:	686	ctctcttctcatcttcat	10206	10225	SEQ ID NO:	1689	atgaagctgtctccaggag	13764	13783	14
SEQ ID NO:	687	tgtcatgtgacactgcag	10226	10245	SEQ ID NO:	1690	ctgcgtaccagaaagaca	12072	12091	14
SEQ ID NO:	688	tgtactgactgcagtacaaa	10232	10251	SEQ ID NO:	1691	tttgagttgccaccatca	11658	11677	14
SEQ ID NO:	689	agctctgtctcgagcaac	10301	10320	SEQ ID NO:	1692	gttgaccacaagcttagct	10539	10558	14
SEQ ID NO:	690	agccgaatccaatttg	10400	10419	SEQ ID NO:	1693	caaagctggcaccagggt	13963	13982	14
SEQ ID NO:	691	ttagaatgaattcaacg	10416	10435	SEQ ID NO:	1694	gcttcaggaaagcttctcaa	13208	13227	14
SEQ ID NO:	692	aaactactgtctcttct	10461	10480	SEQ ID NO:	1695	aggaaggccaagccagttt	12583	12602	14
SEQ ID NO:	693	tacttttcaattgagtc	10575	10594	SEQ ID NO:	1696	atgattatgtaacaagta	12355	12374	14
SEQ ID NO:	694	tcaggctcatgcaagtcag	10910	10929	SEQ ID NO:	1697	ctgacatcttaggcactga	4993	5012	14
SEQ ID NO:	695	atgcaagtcagcccagttc	10918	10937	SEQ ID NO:	1698	gaactcagaagtaggcat	13994	14013	14
SEQ ID NO:	696	tgaatgtaacactaagaa	10975	10994	SEQ ID NO:	1699	ttctcaatttgattttca	8518	8537	14
SEQ ID NO:	697	agaagatcagatggaaaaa	10996	11015	SEQ ID NO:	1700	ttttctaaatggaactct	12165	12184	14
SEQ ID NO:	698	ggctattctattctccatcc	11256	11275	SEQ ID NO:	1701	ggatctaaatgcagtagcc	11624	11643	14
SEQ ID NO:	699	aaagttttggctgataaat	11280	11299	SEQ ID NO:	1702	atttctaaactctctt	9481	9500	14
SEQ ID NO:	700	agttttggctgataaatc	11282	11301	SEQ ID NO:	1703	gaatctggctccctcaact	9039	9058	14
SEQ ID NO:	701	ctgggtgaaactaaatga	11308	11327	SEQ ID NO:	1704	tcattctgggtcttcacg	11027	11046	14
SEQ ID NO:	702	cagagaaatacaaatctat	11405	11424	SEQ ID NO:	1705	atagcatggactctctctg	8865	8884	14
SEQ ID NO:	703	gaggtaaaattccctgaag	11472	11491	SEQ ID NO:	1706	ctctcggttgtaaacctc	12298	12317	14
SEQ ID NO:	704	ctttttgagataaccgtg	11537	11556	SEQ ID NO:	1707	cacggagttactgaaaag	13715	13734	14
SEQ ID NO:	705	gctggaattgtcattctt	11727	11746	SEQ ID NO:	1708	aaggcatctccactcagc	12094	12113	14
SEQ ID NO:	706	gtgtataatgccacttga	11787	11806	SEQ ID NO:	1709	tccaagatgagataaacac	13096	13115	14

271

SEQ ID NO: 735	gctgaggagccgcgcagc	39	58	SEQ ID NO: 738	gctggatgaaccaccagc	11177	11196	23
SEQ ID NO: 736	ctggctgtccaaaagatg	219	238	SEQ ID NO: 739	catcagaaccattgaccag	8126	8145	23
SEQ ID NO: 737	ctgagagttccagtgaggt	283	302	SEQ ID NO: 740	actcaatggtgaaattcag	7457	7476	23
SEQ ID NO: 738	cagtgacccctgaagagg	396	415	SEQ ID NO: 741	cctcacttctcttgagct	8969	8988	23
SEQ ID NO: 739	ctctgaggagttgtctga	464	483	SEQ ID NO: 742	tgcaaaactgacttcagag	11391	11410	23
SEQ ID NO: 740	acatcaagaggggcatcat	574	593	SEQ ID NO: 743	atgagcttcttgagcatgt	7042	7061	23
SEQ ID NO: 741	ctgatcagcagcagccagt	822	841	SEQ ID NO: 744	actggacttctctagtcag	8801	8820	23
SEQ ID NO: 742	ggacgctaagaggaagcat	857	876	SEQ ID NO: 745	atgcctactgttccatgtcc	11346	11365	23
SEQ ID NO: 743	agctgtttgaagactctc	1079	1098	SEQ ID NO: 746	gagaagtgtctcaaaagt	12403	12422	23
SEQ ID NO: 744	tgaaaaaactaacattctc	1105	1124	SEQ ID NO: 747	gagatcaacacaaatctca	13104	13123	23
SEQ ID NO: 745	ctgagctgagaggcctcag	1168	1187	SEQ ID NO: 748	ctgaattactgcacctcag	3027	3046	23
SEQ ID NO: 746	tgaacgtgtgcattgcaa	1303	1322	SEQ ID NO: 749	tgtgtagacgaagggttca	7848	7867	23
SEQ ID NO: 747	cctgtatgcgctgagcca	1432	1451	SEQ ID NO: 750	tggcactgtttggagaagg	9130	9149	23
SEQ ID NO: 748	aggagctgtggacattgc	1492	1511	SEQ ID NO: 751	gcaagtcagccagattctc	10920	10939	23
SEQ ID NO: 749	atttgattctgogggtcat	1567	1586	SEQ ID NO: 752	atgaaccaatgacaaaat	7420	7439	23
SEQ ID NO: 750	tccagaactcaagtctca	1619	1638	SEQ ID NO: 753	tgaatacaatgtctctgga	5512	5531	23
SEQ ID NO: 751	ggttcttctcagacttctc	1736	1755	SEQ ID NO: 754	gaaataccaagtcaaaacc	10447	10466	23
SEQ ID NO: 752	gttgatgaggagtcctca	1802	1821	SEQ ID NO: 755	tgaaaaagctgcaatcaac	13726	13745	23
SEQ ID NO: 753	tccaagatctgaaaagtt	1933	1952	SEQ ID NO: 756	aactgtcttccaaaatgga	3544	3563	23
SEQ ID NO: 754	agltagtgaaagaagtctc	1948	1967	SEQ ID NO: 757	agaattcataatcccaact	8267	8286	23
SEQ ID NO: 755	gaagggaatctatatattg	2076	2095	SEQ ID NO: 758	caaaacctactgtctctc	10459	10478	23
SEQ ID NO: 756	ggaagctcttttgggaag	2213	2232	SEQ ID NO: 759	cttcacataccagaattcc	8316	8335	23
SEQ ID NO: 757	tgaataatgtcagtggt	2366	2385	SEQ ID NO: 760	aacaacacagggcattcca	9648	9667	23
SEQ ID NO: 758	gatttgaatccaaagaag	2400	2419	SEQ ID NO: 761	cttcagtctcctagaaatc	10029	10048	23
SEQ ID NO: 759	tccaagaagtcocggaag	2409	2428	SEQ ID NO: 762	cttcagcctgttcttgga	4943	4962	23
SEQ ID NO: 760	aggaagggtcacaagaatg	2562	2581	SEQ ID NO: 763	catttagagctgccagtcct	10012	10031	23
SEQ ID NO: 761	agaatgactttttctca	2575	2594	SEQ ID NO: 764	tgaagatgacgactttct	12152	12171	23
SEQ ID NO: 762	tttgtacaatatgggca	2757	2776	SEQ ID NO: 765	tgccagttgaaaacaaa	11807	11826	23

273

					NO:								
SEQ ID NO:	791	ggagccgccccaccgcagc	96	115	SEQ ID NO:	1794	gctgtttgaagactctc	1080	1099	13			
SEQ ID NO:	792	gaagaggaaatgctggaaa	192	211	SEQ ID NO:	1795	ttcaagtctcgaccttc	8301	8320	13			
SEQ ID NO:	793	caaaagatgcgacccgatt	229	248	SEQ ID NO:	1796	aaacttatggggatttg	7077	7096	13			
SEQ ID NO:	794	attcaagcacctccggag	245	264	SEQ ID NO:	1797	cttcacatttcaaggat	10059	10078	13			
SEQ ID NO:	795	gttcagtgagctccctgg	289	308	SEQ ID NO:	1798	ccagcaagtcctcgagaac	8602	8621	13			
SEQ ID NO:	796	gactgctgatcaagaagt	308	327	SEQ ID NO:	1799	acttgaagaaaagatagtc	13316	13335	13			
SEQ ID NO:	797	gtgccaccaggatcaactg	325	344	SEQ ID NO:	1800	cagtgaagctgcaggccac	10696	10715	13			
SEQ ID NO:	798	gatcaactgcaaggttgag	335	354	SEQ ID NO:	1801	ctcacctccacctctgatc	4740	4759	13			
SEQ ID NO:	799	actgcaaggttgagctgga	340	359	SEQ ID NO:	1802	tcactcacatcctccagt	1281	1300	13			
SEQ ID NO:	800	ccagctctgcagctctac	365	384	SEQ ID NO:	1803	gatgtgtcacctacactgg	1335	1354	13			
SEQ ID NO:	801	agcttcatctgaagacca	375	394	SEQ ID NO:	1804	tgtgtgctgagaatgagct	5104	5123	13			
SEQ ID NO:	802	cttcatcctgaagaccagc	377	396	SEQ ID NO:	1805	gctggagtaaaactggaag	2688	2707	13			
SEQ ID NO:	803	ccagccagtcacactgaa	391	410	SEQ ID NO:	1806	ttcaagatgactgcactgg	1531	1550	13			
SEQ ID NO:	804	cagtcacacctgaagagg	396	415	SEQ ID NO:	1807	cctcacagagctatcactg	5222	5241	13			
SEQ ID NO:	805	tggcttaacctgagggc	419	438	SEQ ID NO:	1808	gccactggtgcgctgcca	3525	3544	13			
SEQ ID NO:	806	cttcaacctgagggcaaa	422	441	SEQ ID NO:	1809	tttagccacaacttggag	2199	2218	13			
SEQ ID NO:	807	ttcaacctgagggcaag	423	442	SEQ ID NO:	1810	cttgacaggcatittgaa	9719	9738	13			
SEQ ID NO:	808	cttgctgaagaaaaccaag	443	462	SEQ ID NO:	1811	ctgaaattcaatcacaa	9066	9085	13			
SEQ ID NO:	809	tgctgaagaaaaccaagaa	445	464	SEQ ID NO:	1812	ttctgtcgcctatcagca	5639	5658	13			
SEQ ID NO:	810	ttgctgcagccatgtccag	475	494	SEQ ID NO:	1813	ctggtcagtttgcagcaa	2996	3015	13			
SEQ ID NO:	811	tgctgcagccatgtccagg	476	495	SEQ ID NO:	1814	cctggtcagtttgcagcaa	2995	3014	13			
SEQ ID NO:	812	agccatgctcaggtatgag	482	501	SEQ ID NO:	1815	ctcacatcctccagtggt	1285	1304	13			
SEQ ID NO:	813	agctcaagctggccattcc	499	518	SEQ ID NO:	1816	ggaactacacaaaaagct	7481	7500	13			
SEQ ID NO:	814	agaagggaagcaggttttc	518	537	SEQ ID NO:	1817	gaaatcttcaattattct	13813	13832	13			
SEQ ID NO:	815	aagggaagcaggttttct	520	539	SEQ ID NO:	1818	aggacacaaaataacctt	7564	7583	13			
SEQ ID NO:	816	agaagatgaacctactta	547	566	SEQ ID NO:	1819	taagaacttgcactctt	4844	4863	13			
SEQ ID NO:	817	atcctgaacatcaagagg	567	586	SEQ ID NO:	1820	ccctaacagatttgaggat	7969	7988	13			
SEQ ID NO:	818	tctgaacatcaagagg	568	587	SEQ ID NO:	1821	ccctaacagatttgaggat	7968	7987	13			

				NO:									
SEQ ID NO:	819	ctgaacatcaagaggggca	570	589	SEQ ID NO:	1822	tgccgtgccttgaagtcag	7900	7919	13			
SEQ ID NO:	820	aacatcaagaggggcatca	573	592	SEQ ID NO:	1823	tgataaaaaccaagatgt	6290	6309	13			
SEQ ID NO:	821	acatcaagaggggcatcat	574	593	SEQ ID NO:	1824	atgataaaaaccaagatgt	6289	6308	13			
SEQ ID NO:	822	tcatctctgocctcctgt	589	608	SEQ ID NO:	1825	accaccagttttagatga	7405	7424	13			
SEQ ID NO:	823	ttcccccagagacagaaga	607	626	SEQ ID NO:	1826	tctccacatttcaaggaa	10058	10077	13			
SEQ ID NO:	824	gaagaagccaagcaaggt	621	640	SEQ ID NO:	1827	acaccttccacattcctc	8071	8090	13			
SEQ ID NO:	825	ttgttcttggtatccgt	639	658	SEQ ID NO:	1828	acactaatacttccacaa	8767	8786	13			
SEQ ID NO:	826	tgatggaaactgctccac	655	674	SEQ ID NO:	1829	gtgagggaacacattaca	2920	2939	13			
SEQ ID NO:	827	aaactgtccactcacttt	662	681	SEQ ID NO:	1830	aaagaacacagcattgttt	4532	4551	13			
SEQ ID NO:	828	actcatttaccgtcaaga	672	691	SEQ ID NO:	1831	tcttactttccattgagt	10572	10591	13			
SEQ ID NO:	829	ctttaccgtcaagacagg	677	696	SEQ ID NO:	1832	cctccagctctctgggaaag	2483	2502	13			
SEQ ID NO:	830	ttaccgtcaagacaggaa	679	698	SEQ ID NO:	1833	ttctaaagctggatgtaa	11169	11188	13			
SEQ ID NO:	831	acgaggaagggcaatgtgg	690	709	SEQ ID NO:	1834	ccacaagtcacatctcgt	5956	5975	13			
SEQ ID NO:	832	cgaggaagggcaatgtgac	691	710	SEQ ID NO:	1835	gccagaagtggagatctcg	3507	3526	13			
SEQ ID NO:	833	gaggaagggcaatgtggca	692	711	SEQ ID NO:	1836	tgccagctctccatgacctc	2468	2487	13			
SEQ ID NO:	834	ggaagggcaatgtggcaac	694	713	SEQ ID NO:	1837	gttgccttaaggadtcc	13356	13375	13			
SEQ ID NO:	835	gaaggccaatgtggcaaca	695	714	SEQ ID NO:	1838	tggtgatgaggagctcctc	1801	1820	13			
SEQ ID NO:	836	caggcatcagcccacttgc	769	788	SEQ ID NO:	1839	gcaagctcttctggccctg	3011	3030	13			
SEQ ID NO:	837	aggcatcagcccacttgc	770	789	SEQ ID NO:	1840	agcaagctcttctggcct	3010	3029	13			
SEQ ID NO:	838	tcagcccacttgcctcat	775	794	SEQ ID NO:	1841	atgaagtcacagcatctga	12680	12679	13			
SEQ ID NO:	839	gtcaactctgatcagcagc	815	834	SEQ ID NO:	1842	gctgactttaaatctgac	4811	4830	13			
SEQ ID NO:	840	ggagcctaaggaagcat	857	876	SEQ ID NO:	1843	atgcactgtttctgagctc	9331	9350	13			
SEQ ID NO:	841	aaggagcaacacctcttcc	894	913	SEQ ID NO:	1844	ggaatatcttagcatcctt	13457	13476	13			
SEQ ID NO:	842	aggagcaacacctcttcc	895	914	SEQ ID NO:	1845	aggaatatcttagcatcct	13456	13475	13			
SEQ ID NO:	843	caacacctcttctgcctt	900	919	SEQ ID NO:	1846	aaggctgactctggtgtg	4284	4303	13			
SEQ ID NO:	844	aacacctcttctgcctt	901	920	SEQ ID NO:	1847	aaagcaggccgaagctgtt	1067	1086	13			
SEQ ID NO:	845	acaagaataagtatgggat	925	944	SEQ ID NO:	1848	atccatgatctacattgt	6786	6805	13			
SEQ ID NO:	846	caagaataagtatgggat	926	945	SEQ ID NO:	1849	catcacttcaagacctg	1238	1257	13			

WU 2004/09/15				TC 2004/09/15			
SEQ ID NO:				NO:			
SEQ ID NO: 847	tagcacaagtgacacagac	946	965	SEQ ID NO: 1850	gtctcttgcttatgtcta	4584	4603
SEQ ID NO: 848	agcacaagtgacacagact	947	966	SEQ ID NO: 1851	agtcctcttgcttatgtct	4583	4602
SEQ ID NO: 849	gcacaagtgacacagactt	948	967	SEQ ID NO: 1852	aagtgtagctctctggtgc	5091	5110
SEQ ID NO: 850	aacttgaagacacacacaaa	970	989	SEQ ID NO: 1853	tttgaggattccatcagtt	7979	7998
SEQ ID NO: 851	gctctttggtgaaggtaac	1000	1019	SEQ ID NO: 1854	gtacctacttttggaacgc	8364	8383
SEQ ID NO: 852	ctttggtagaggtactaag	1004	1023	SEQ ID NO: 1855	cttatgggatttctaagag	11159	11178
SEQ ID NO: 853	tactaagaagatgggacct	1016	1035	SEQ ID NO: 1856	gagggtagcataacagta	10329	10348
SEQ ID NO: 854	tttgagagcaccaaatcca	1038	1057	SEQ ID NO: 1857	tggaggtgcagtggaacaa	10372	10391
SEQ ID NO: 855	agagcaccaaatccacatc	1042	1061	SEQ ID NO: 1858	gatggatatgacctctct	4868	4887
SEQ ID NO: 856	agctgtttgaagactctc	1079	1098	SEQ ID NO: 1859	gagaacatactgggcagct	5872	5891
SEQ ID NO: 857	tgaaaaaactaaccatctc	1105	1124	SEQ ID NO: 1860	gagaaatcaatgccttca	7104	7123
SEQ ID NO: 858	gaaaaaactaaccatctct	1106	1125	SEQ ID NO: 1861	agagccaggctcgagcttc	11044	11063
SEQ ID NO: 859	tctgagcaaaatatccaga	1122	1141	SEQ ID NO: 1862	tctgatgaggaaactcaga	12252	12271
SEQ ID NO: 860	tctcttaataagctggtt	1148	1167	SEQ ID NO: 1863	aacctccactttttlaga	6318	6337
SEQ ID NO: 861	ctgagctgagaggccctcag	1168	1187	SEQ ID NO: 1864	ctgatcccgagccctcag	1359	1378
SEQ ID NO: 862	tgaagcagtcacatctctc	1190	1209	SEQ ID NO: 1865	gagaaaaatcaatgccttca	7104	7123
SEQ ID NO: 863	aagcagtcacatctctctt	1192	1211	SEQ ID NO: 1866	aagaggcagctcttggtt	12289	12308
SEQ ID NO: 864	ctctcttccacagctgat	1204	1223	SEQ ID NO: 1867	atcaaaagaagcccaagag	12938	12957
SEQ ID NO: 865	tcttgccacagctgattga	1207	1226	SEQ ID NO: 1868	tcaagtaatttgggaaga	12271	12290
SEQ ID NO: 866	cttgccacagctgattgag	1208	1227	SEQ ID NO: 1869	ctcaatttttgatttcaag	8520	8539
SEQ ID NO: 867	tgaggtgtccagcccatc	1223	1242	SEQ ID NO: 1870	gatggaacctctccctca	4725	4744
SEQ ID NO: 868	tcagttggacagcctcag	1259	1278	SEQ ID NO: 1871	ctgacatctaggcacctga	4993	5012
SEQ ID NO: 869	acatctctccagtggtgaa	1288	1307	SEQ ID NO: 1872	ttcagaagctaagcaatgt	7231	7250
SEQ ID NO: 870	gcacagcagctgcgagaga	1377	1396	SEQ ID NO: 1873	tctctgaagacaacgtgc	12315	12334
SEQ ID NO: 871	cagcagctgcgagagatct	1380	1399	SEQ ID NO: 1874	agataacattaaacagctg	13043	13062
SEQ ID NO: 872	gcgagggatcagcgcagcc	1407	1426	SEQ ID NO: 1875	ggctcaacacagacatcgc	5710	5729
SEQ ID NO: 873	aagacaacacctacaggga	1470	1489	SEQ ID NO: 1876	tccagaaaacctctctt	3928	3947
SEQ ID NO: 874	caggagctctgtggacattg	1491	1510	SEQ ID NO: 1877	caatggagagtccaacctg	4652	4671

				NO:					
SEQ ID NO: 875	aggagctgctggacattgc	1492	1511	SEQ ID NO:	1878	gcaagggttcactgttct	7856	7875	13
SEQ ID NO: 876	ctgctggacattgctaatt	1497	1516	SEQ ID NO:	1879	aattgggaagaagaggcag	12279	12298	13
SEQ ID NO: 877	gattacacctattgattc	1557	1576	SEQ ID NO:	1880	gaatattttgagaggaatc	6345	6364	13
SEQ ID NO: 878	atttgattctcgggtcat	1567	1586	SEQ ID NO:	1881	atgaagtagaccaacaat	7153	7172	13
SEQ ID NO: 879	ctcggggtcattggaat	1574	1593	SEQ ID NO:	1882	atttgaagaataacaga	6428	6447	13
SEQ ID NO: 880	aacctggagcagttaaat	1601	1620	SEQ ID NO:	1883	agttctccactctaggt	9954	9973	13
SEQ ID NO: 881	ggagcagttactocagaa	1607	1626	SEQ ID NO:	1884	ttctgaaatccaatctcc	8392	8411	13
SEQ ID NO: 882	actocagaactcaagttct	1617	1636	SEQ ID NO:	1885	aagatcgacagcttgagt	11646	11665	13
SEQ ID NO: 883	tocagaactcaagttcca	1619	1638	SEQ ID NO:	1886	tgaactcagaagaattgga	1912	1931	13
SEQ ID NO: 884	aagtaacaagccatcactg	1655	1674	SEQ ID NO:	1887	cagtcactgtagaaaactt	4421	4440	13
SEQ ID NO: 885	gccatcactgatgatccag	1664	1683	SEQ ID NO:	1888	ctggaactctctccatggc	10875	10894	13
SEQ ID NO: 886	ccatcactgatgatccaga	1665	1684	SEQ ID NO:	1889	ctggaactcagaaggatgg	13991	14010	13
SEQ ID NO: 887	atccagaagaagcgcatcc	1677	1696	SEQ ID NO:	1890	ggatttccaaagctggat	11165	11184	13
SEQ ID NO: 888	cagaaagctgccatccagg	1680	1699	SEQ ID NO:	1891	cctgaatacaatgctctg	5510	5529	13
SEQ ID NO: 889	acaaggaccaggaggttct	1723	1742	SEQ ID NO:	1892	agaacagcatttgtttgt	4534	4553	13
SEQ ID NO: 890	aggaccaggaggttctct	1726	1745	SEQ ID NO:	1893	agaagctaagcaatgtcct	7234	7253	13
SEQ ID NO: 891	accaggaggttctcttcca	1729	1748	SEQ ID NO:	1894	tgaaggctgactctgtgt	4282	4301	13
SEQ ID NO: 892	tctcagacttctctgat	1742	1761	SEQ ID NO:	1895	atcaggaagggtcacaaga	2559	2578	13
SEQ ID NO: 893	ttcagacttctctgatga	1744	1763	SEQ ID NO:	1896	tcattactctctgggtgaa	11299	11318	13
SEQ ID NO: 894	gttgatgaggagctctcca	1802	1821	SEQ ID NO:	1897	tgaatctgctccctcaac	9038	9057	13
SEQ ID NO: 895	cttcacaggcagataataa	1816	1835	SEQ ID NO:	1898	ttaatcgagaggtatgaag	7140	7159	13
SEQ ID NO: 896	ttcacaggcagatattaac	1817	1836	SEQ ID NO:	1899	gttaatcgagaggtatgaa	7139	7158	13
SEQ ID NO: 897	ggcagatattaacaaat	1823	1842	SEQ ID NO:	1900	aattgcattagatgatgcc	6581	6600	13
SEQ ID NO: 898	atattaacaaaattgtcca	1828	1847	SEQ ID NO:	1901	tggagttgtgacaataat	2752	2771	13
SEQ ID NO: 899	acaaaattgtccaaattct	1834	1853	SEQ ID NO:	1902	agaacacagcatttgtttg	4534	4553	13
SEQ ID NO: 900	gagcaagtgaagaacttg	1869	1888	SEQ ID NO:	1903	caaatgacatgatgggtc	5326	5345	13
SEQ ID NO: 901	gtgaagaacttgggctt	1875	1894	SEQ ID NO:	1904	aagcatctgatgatacac	12669	12688	13
SEQ ID NO: 902	agaacttgtggcttccca	1879	1898	SEQ ID NO:	1905	tgggctctgccagattct	8901	8920	13

SEQ ID NO:	903	ttgtggcttcccatattg	1884	1903	SEQ ID NO:	1906	caataagatcaatagcaaa	8990	9009	13
SEQ ID NO:	904	tggtctcccatattgcca	1888	1907	SEQ ID NO:	1907	ttggctcacatgaaggcca	7623	7642	13
SEQ ID NO:	905	ttcccatattgccaatc	1882	1911	SEQ ID NO:	1908	gatatacactaggaggaa	12737	12756	13
SEQ ID NO:	906	tcccatattgccaatct	1893	1912	SEQ ID NO:	1909	agatcaaagtaattggga	12268	12287	13
SEQ ID NO:	907	ttgccaatcttgaactc	1900	1919	SEQ ID NO:	1910	gagtcacagtcacagcaa	9344	9363	13
SEQ ID NO:	908	ttgcatatccaagatctga	1926	1945	SEQ ID NO:	1911	tcagtataagtacaccaa	9392	9411	13
SEQ ID NO:	909	tccaagatctgaaaaagtt	1933	1952	SEQ ID NO:	1912	aactccaactgtcatgga	1978	1997	13
SEQ ID NO:	910	ctgaaaaagttagtgaag	1941	1960	SEQ ID NO:	1913	ctttgaagcagtcctcag	7907	7926	13
SEQ ID NO:	911	agtagtgaaagaagttct	1948	1967	SEQ ID NO:	1914	agaatctcaacttccaact	1970	1989	13
SEQ ID NO:	912	aactccaactccaactgt	1972	1991	SEQ ID NO:	1915	acaggggtccttatgatt	12342	12361	13
SEQ ID NO:	913	gicagtgaactcagaaat	1989	2008	SEQ ID NO:	1916	atttgaagaataaatgac	7028	7047	13
SEQ ID NO:	914	tcaactctacaatctgtt	2021	2040	SEQ ID NO:	1917	aacacattgagctattga	6970	6989	13
SEQ ID NO:	915	aactctcaaatctgttct	2023	2042	SEQ ID NO:	1918	gaaaaaggggattgaagtt	10276	10295	13
SEQ ID NO:	916	aaatagaagggaaatcttat	2071	2090	SEQ ID NO:	1919	ataagcaaatgttaattt	5449	5468	13
SEQ ID NO:	917	agaagggaaatcttatatt	2075	2094	SEQ ID NO:	1920	aaatgactgtctgcgttct	4892	4911	13
SEQ ID NO:	918	gaagggaaatcttatattg	2076	2095	SEQ ID NO:	1921	caaaaacatttcaacttc	5279	5298	13
SEQ ID NO:	919	tgtaccaataactacactt	2093	2112	SEQ ID NO:	1922	aaggagaagaagaaaatca	3453	3472	13
SEQ ID NO:	920	tggaattgtctcagctgac	2150	2169	SEQ ID NO:	1923	gtcagccagttccttcca	10924	10943	13
SEQ ID NO:	921	ttgtctcagctgaactca	2154	2173	SEQ ID NO:	1924	tgaggaaactcagatcaaa	12257	12276	13
SEQ ID NO:	922	cttggaggaaaaaggcttt	2183	2202	SEQ ID NO:	1925	aaagcattggtagagcaag	7842	7861	13
SEQ ID NO:	923	tggaaggaaaaggctttga	2185	2204	SEQ ID NO:	1926	tcaagtctgtggattcca	4078	4097	13
SEQ ID NO:	924	ggctttgagccaacattgg	2196	2215	SEQ ID NO:	1927	ccaagaggtatttaaagcc	12950	12969	13
SEQ ID NO:	925	tgagccaacattggaagct	2201	2220	SEQ ID NO:	1928	agcttctgcacactgctca	13513	13532	13
SEQ ID NO:	926	gagccaacattggaagctc	2202	2221	SEQ ID NO:	1929	gagcttctgccactgctc	13512	13531	13
SEQ ID NO:	927	aacattggaagctctttt	2207	2226	SEQ ID NO:	1930	aaaagaacagcatttgtt	4531	4550	13
SEQ ID NO:	928	tggaagctcttttgggaa	2212	2231	SEQ ID NO:	1931	ttccggcacgtgggtcca	3777	3796	13
SEQ ID NO:	929	ctcttttgggaagcaagg	2218	2237	SEQ ID NO:	1932	ccttactgactgtcagag	7790	7809	13
SEQ ID NO:	930	ttttgggaagcaaggatt	2221	2240	SEQ ID NO:	1933	aatcattgaaataaaaa	6722	6741	13

WU 2004/091515				TC70520491515			
SEQ ID NO:			NO:				
931	ttttccagcagcagtgtcaa	2239	2258	934	tgtatgaatcattgtaaaa	6715	6734
932	tggctataccaaagatga	2323	2342	935	tcattgctccggagccaa	2668	2687
933	ataccaaagatgataaaca	2329	2348	936	tgtgtgtttgtaaagtat	6272	6291
934	gagcaggatatgttaaatg	2349	2368	937	caattcagccttgggctc	4254	4273
935	alggtaaatggaataatgc	2358	2377	938	gcattgctagtgttccat	9946	9965
936	tggtaaatggaataatgct	2359	2378	939	agcacagctacgaaaaacca	10801	10820
937	taaatggaataatgctcag	2362	2381	940	ctgaaagagatgaaatta	13059	13078
938	tggaaataatgctcagtgt	2366	2385	941	aacagatttgaggattcca	7973	7992
939	tcagtggtgagaagctgat	2377	2396	942	atcacaaactctccactga	9534	9553
940	cagtggtgagaagctgatt	2378	2397	943	aatcacaaactctccactg	9533	9552
941	agtggtgagaagctgatta	2379	2398	944	taatcacaaactctccact	9532	9551
942	gattaagaattgaaatcc	2393	2412	945	ggatactaagtaaccaatc	6866	6885
943	gatttgaatccaagaag	2400	2419	946	cttcgtttaccagaaatc	8240	8259
944	atttgaatccaagaagt	2401	2420	947	acttcggtttaccagaaat	8239	8258
945	atccaaagaagctccggaa	2408	2427	948	ttccaatttccctgtggat	3680	3699
946	tccaagaagctccgggaag	2409	2428	949	cttccaatttccctgtgga	3679	3698
947	agagcctaccctccgatct	2430	2449	950	agattaatccgtcggtct	8563	8582
948	gagcctaccctccgatctt	2431	2450	951	aagattaatccgtcggtct	8562	8581
949	cttgggagaggagctgtgt	2447	2466	952	accactgggacctaccaag	12519	12538
950	ggagctgtgtttgccagt	2456	2475	953	actggtggcaaaacctcc	2726	2745
951	tgtgtttgccagctcca	2461	2480	954	tggagaagccacctccaa	10763	10782
952	cagctctcatgacctccag	2471	2490	955	ctggctgcctgccaaactg	3530	3549
953	ctccatgacctccagctcc	2475	2494	956	ggagtcattgctccggag	2664	2683
954	ctgggaaagctgctctga	2483	2512	957	tcgaaagctacctccag	7931	7950
955	gaggtcatcaggaagggt	2553	2572	958	agccagaagtgcagctct	3506	3525
956	aagaatgactttttcttc	2574	2593	959	gaaggcatctgggagctt	3827	3846
957	ctttttctctacacatc	2582	2601	960	gatgctacaacactaaag	6099	6118
958	catctcatcgagaatgcc	2597	2616	961	ggcacttccaaaattgatg	10710	10729

SEQ ID NO:	959	cttcatggagaatgcttt	2600	2619	SEQ ID NO:	1962	aaagttaattgggaagaag	12273	12292	13
SEQ ID NO:	960	aatgctttgaactcccca	2610	2629	SEQ ID NO:	1963	tggctgtggtcagccatt	5729	5748	13
SEQ ID NO:	961	gccttgaactccccactg	2613	2632	SEQ ID NO:	1964	cagctgaacatgcaggc	5375	5394	13
SEQ ID NO:	962	caaggctggagtaaaactg	2684	2703	SEQ ID NO:	1965	cagtgcaacgaccaactg	5072	5091	13
SEQ ID NO:	963	tggagtaaaactggaagta	2690	2709	SEQ ID NO:	1966	tactccaagcagctcca	3051	3070	13
SEQ ID NO:	964	ggaagtagccaacatgcag	2702	2721	SEQ ID NO:	1967	ctgcatctcagagatgcc	4098	4117	13
SEQ ID NO:	965	tttgtacaaatatgggca	2757	2776	SEQ ID NO:	1968	tgccttgtgtacacaaa	11228	11247	13
SEQ ID NO:	966	tgtgacaaatatgggcatc	2759	2778	SEQ ID NO:	1969	gatgggtctctacgccaca	4377	4396	13
SEQ ID NO:	967	ggacttcgttaggagtggtg	2786	2805	SEQ ID NO:	1970	cccaaggccacaggggtcc	12333	12352	13
SEQ ID NO:	968	gtgggtccagatgaacac	2800	2819	SEQ ID NO:	1971	gtgtctagacctctcac	4171	4190	13
SEQ ID NO:	969	ttcacgagtcgggtctgg	2826	2845	SEQ ID NO:	1972	ccagaatctgaccaggaa	12554	12573	13
SEQ ID NO:	970	agtgggtctggaggctca	2833	2852	SEQ ID NO:	1973	tgagaactacgagctgact	4799	4818	13
SEQ ID NO:	971	tgggtctgtagggctcatg	2835	2854	SEQ ID NO:	1974	catgaaggccaaattccga	7631	7650	13
SEQ ID NO:	972	aaaagctgggaagctgaag	2861	2880	SEQ ID NO:	1975	cttcagacacctgattt	7943	7962	13
SEQ ID NO:	973	aagctgaagtttatcttc	2871	2890	SEQ ID NO:	1976	gaattacaattgttgctt	6261	6280	13
SEQ ID NO:	974	gagaccagtcagctctc	2900	2919	SEQ ID NO:	1977	gagctcaggaagctctc	13206	13225	13
SEQ ID NO:	975	gcaacacattacatttgtt	2926	2945	SEQ ID NO:	1978	accagtcagatattgttc	10183	10202	13
SEQ ID NO:	976	acattacatttggtctcta	2931	2950	SEQ ID NO:	1979	tagaataatgaactaaatgt	11881	11900	13
SEQ ID NO:	977	cattacatttggtctctac	2932	2951	SEQ ID NO:	1980	gtagctgagaaatcaatg	7098	7117	13
SEQ ID NO:	978	aaacggaggtgatccacc	2956	2975	SEQ ID NO:	1981	gggtgataccctgaagttt	3197	3216	13
SEQ ID NO:	979	attgagaacaggcagctct	2979	2998	SEQ ID NO:	1982	aggaaaagcgcaacctaat	12023	12042	13
SEQ ID NO:	980	tgaacacaggcagctctgg	2981	3000	SEQ ID NO:	1983	ccagcttccccatctca	8333	8352	13
SEQ ID NO:	981	ctgcacctcaggcgcttac	3035	3054	SEQ ID NO:	1984	gtaagaaaatacagagcag	6432	6451	13
SEQ ID NO:	982	tcacagactccgcctct	3066	3085	SEQ ID NO:	1985	aggacagagccttggtgga	3184	3203	13
SEQ ID NO:	983	ctgacggggacaccagat	3093	3112	SEQ ID NO:	1986	atctgatgaggaaactcag	12251	12270	13
SEQ ID NO:	984	tagagctggaaactgaggcc	3112	3131	SEQ ID NO:	1987	ggcctctctggggcactcta	5136	5155	13
SEQ ID NO:	985	ctatgagctccagagagag	3167	3186	SEQ ID NO:	1988	ctctcacaataaaglatag	6541	6560	13
SEQ ID NO:	986	cttggtgataccctgaag	3194	3213	SEQ ID NO:	1989	cttcagggaagcttccaag	13209	13228	13

281

SEQ ID NO: 2026	gaaggctgacctggtgt	4283	4302	SEQ ID NO: 2332	aacctatgccttaattctc	13161	13180	13
SEQ ID NO: 2027	tggtgtgacctgtcttc	4295	4314	SEQ ID NO: 2333	ggaagaagttaaacacaca	6957	6976	13
SEQ ID NO: 2028	ctcgtcttcacaaatgtg	4304	4323	SEQ ID NO: 2334	acaccttgacatgcagg	11080	11099	13
SEQ ID NO: 2029	ctcgtcttcacaaatgtg	4305	4324	SEQ ID NO: 2335	gcacacctgacatgcag	11079	11098	13
SEQ ID NO: 2030	ctcacaatgtgcaagatg	4311	4330	SEQ ID NO: 2336	atcgcgtgcgtcgaagga	8569	8588	13
SEQ ID NO: 2031	atgaccacaagaataatcgt	4344	4363	SEQ ID NO: 2337	acgtcgtgtgccttcata	9976	9995	13
SEQ ID NO: 2032	atgaccacaagaataatcgt	4345	4364	SEQ ID NO: 2338	gacgtcgtgtgccttcata	9975	9994	13
SEQ ID NO: 2033	gaatacgtctacataatca	4355	4374	SEQ ID NO: 2339	gatattctgaattatcttc	6479	6498	13
SEQ ID NO: 2034	ttctagatctgaataatca	4398	4417	SEQ ID NO: 2340	gtattacatgatttgaaa	6677	6696	13
SEQ ID NO: 2035	gattctgaataatacaattca	4404	4423	SEQ ID NO: 2341	gaagtagctgagaaaatc	7094	7113	13
SEQ ID NO: 2036	gaacaacacccagtctcaaa	4441	4460	SEQ ID NO: 2342	ttgaaaaattctcttttc	9206	9225	13
SEQ ID NO: 2037	ccagctctcaaaagggttta	4448	4467	SEQ ID NO: 2343	taaatcttaactctctggg	11294	11313	13
SEQ ID NO: 2038	ctcaaaagggtttactaata	4454	4473	SEQ ID NO: 2344	tattcaaaaactgagttgag	12223	12242	13
SEQ ID NO: 2039	ctcaaaagggtttactaata	4455	4474	SEQ ID NO: 2345	tattcaaaaactgagttgag	12222	12241	13
SEQ ID NO: 2040	aaaagggtttactaataatc	4457	4476	SEQ ID NO: 2346	gaatttgaaagtctgtttt	9272	9291	13
SEQ ID NO: 2041	gaacacagatctgtgtgtg	4535	4554	SEQ ID NO: 2347	gacagcatctctgtgtgtg	11206	11225	13
SEQ ID NO: 2042	attgtgtgttcaagaagat	4543	4562	SEQ ID NO: 2348	actataaaaaatataaaaat	8014	8033	13
SEQ ID NO: 2043	caagattgtggtggcagtt	4561	4580	SEQ ID NO: 2349	aactctcaagctcaagttga	13414	13433	13
SEQ ID NO: 2044	ttcagagctctctgtgtct	4578	4597	SEQ ID NO: 2350	agaagatggcaaaattgaa	11987	12006	13
SEQ ID NO: 2045	cagagctctctgtgtctat	4580	4599	SEQ ID NO: 2351	atagcatgactctctgtct	8865	8884	13
SEQ ID NO: 2046	atgctcaaaaggccatattg	4597	4616	SEQ ID NO: 2352	catlltgagatcaagcgcat	9237	9256	13
SEQ ID NO: 2047	gacatatggcctgtctgtg	4606	4625	SEQ ID NO: 2353	caagtgtgcgaagtgatgc	9364	9383	13
SEQ ID NO: 2048	gagctcaaacctgaggttta	4659	4678	SEQ ID NO: 2354	taaaagtgcacatttactc	6182	6201	13
SEQ ID NO: 2049	agctcaaacctgaggttttaa	4660	4679	SEQ ID NO: 2355	ttacacaggaagatagact	9300	9319	13
SEQ ID NO: 2050	ctcactctcaaaaggccacaa	4684	4703	SEQ ID NO: 2356	gtcaggttaagtgctclagg	9368	9387	13
SEQ ID NO: 2051	gaagatgggaacctctctccc	4722	4741	SEQ ID NO: 2357	gggaagaagaggcagcttc	12283	12302	13
SEQ ID NO: 2052	gatctgcgaagtggcatc	4754	4773	SEQ ID NO: 2358	gatgaggaactacatgata	12255	12274	13
SEQ ID NO: 2053	gactctgcaaaagtgcataca	4755	4774	SEQ ID NO: 2359	gatgaggaactacatgata	12254	12273	13
SEQ ID NO: 2054	gcttctcctaagaatgatga	4785	4804	SEQ ID NO: 2360	ttcgtgtgtctagaaaagc	5969	5988	13
SEQ ID NO: 2055	ggtatgagaactacagcgtg	4796	4815	SEQ ID NO: 2361	cagcttaagaagcacatac	6912	6931	13
SEQ ID NO: 2056	ctacaagaatgagataga	4860	4879	SEQ ID NO: 2362	catllttccacaataataga	13024	13043	13
SEQ ID NO: 2057	ctcgtcgtcttgaaatca	4899	4918	SEQ ID NO: 2363	gatacaagaaaaactgcag	6893	6912	13
SEQ ID NO: 2058	caatgaggtttcttcagcgc	4932	4951	SEQ ID NO: 2364	ggctctatgtctgtaaatga	5340	5359	13
SEQ ID NO: 2059	ttctgtgcatcaataattcc	4955	4974	SEQ ID NO: 2365	ggaaggacaaaggccacgaa	12541	12560	13
SEQ ID NO: 2060	ccatgcttcttgagtttaaat	4973	4992	SEQ ID NO: 2366	atttllattctctgcacatgg	10095	10114	13
SEQ ID NO: 2061	ttctaggcactgcacaaat	4999	5018	SEQ ID NO: 2367	attttttgcaagtttaaga	14011	14030	13
SEQ ID NO: 2062	caaaaggcgacactaaagat	5032	5051	SEQ ID NO: 2368	attcatgatctacatttgt	6786	6805	13
SEQ ID NO: 2063	tgcaacgacacacttgaaag	5075	5094	SEQ ID NO: 2369	cttgagtggaacacaaatgca	5177	5196	13
SEQ ID NO: 2064	caacttgagtgatgtctgc	5084	5103	SEQ ID NO: 2370	gagatgagagatgcogttg	6231	6250	13
SEQ ID NO: 2065	gctgagaaatgagctgaat	5108	5127	SEQ ID NO: 2371	ttatctcttlltllctcagc	9214	9233	13
SEQ ID NO: 2066	gagagctgtggcctctctg	5127	5146	SEQ ID NO: 2372	gacatacagaagaaaactgc	6491	6510	13
SEQ ID NO: 2067	ttctctggggcatctatgaa	5140	5159	SEQ ID NO: 2373	ttcatctaaattgggagaga	6491	6510	13
SEQ ID NO: 2068	ctgggggcatctatgaaat	5142	5161	SEQ ID NO: 2374	atttttgagaataatcaga	6428	6447	13
SEQ ID NO: 2069	aacacaaatgcacaaatcag	5185	5204	SEQ ID NO: 2375	ctgaagcattaaaactgtt	7498	7517	13
SEQ ID NO: 2070	ctcacagagctatcatctgg	5223	5242	SEQ ID NO: 2376	ccagatgctgcacagtgag	8141	8160	13
SEQ ID NO: 2071	tggaagtgctattcaggc	5239	5258	SEQ ID NO: 2377	gctcatgcttccatgtccca	11348	11367	13
SEQ ID NO: 2072	ttcaaggtcagtcagaaga	5295	5314	SEQ ID NO: 2378	ctcagtgacgaatagtaa	11969	11988	13
SEQ ID NO: 2073	aatgacatgatggcgctcat	5328	5347	SEQ ID NO: 2379	gatattatctgaattatct	6478	6497	13
SEQ ID NO: 2074	gctcatgctctgaaatgaa	5341	5360	SEQ ID NO: 2380	ttcagcactcatgacagc	5738	5757	13
SEQ ID NO: 2075	atatgctgaaatgaaattt	5345	5364	SEQ ID NO: 2381	aaatgactatgtctaata	6694	6713	13
SEQ ID NO: 2076	ctgaacatctcagcgctta	5378	5397	SEQ ID NO: 2382	taagaaccagaagatcaga	10988	11007	13
SEQ ID NO: 2077	gaacattgcaggctcttca	5381	5400	SEQ ID NO: 2383	gtatctgcagctgaggttc	12482	12501	13

SEQ ID NO: 2078	gcaggctatcacctggac	5387	5406	SEQ ID NO: 2384	gtctggattccacatgca	11844	11863	13
SEQ ID NO: 2079	caaaacitgacacattt	5412	5431	SEQ ID NO: 2385	aaatctctgacattgtga	7362	7381	13
SEQ ID NO: 2080	atttaccagctctgcaagt	5427	5446	SEQ ID NO: 2386	acttaaaaaataaaaaat	8014	8033	13
SEQ ID NO: 2081	ctctgacagatttataag	5435	5454	SEQ ID NO: 2387	cttactgaaatccaagag	10668	10685	13
SEQ ID NO: 2082	gttaatttaccagctacagc	5460	5479	SEQ ID NO: 2388	gctgcatgtgctggaac	5570	5589	13
SEQ ID NO: 2083	tctctgttgtaactcttta	5483	5502	SEQ ID NO: 2389	taaaagattacttggaaa	7267	7286	13
SEQ ID NO: 2084	ccctaaaggagcctaccaa	5588	5607	SEQ ID NO: 2390	tgccagtaagtgtcagg	9368	9387	13
SEQ ID NO: 2085	aaaaggagcctaccaaaat	5591	5610	SEQ ID NO: 2391	atttacaattgtgtttctt	6263	6282	13
SEQ ID NO: 2086	aggagcctaccaaaataat	5594	5613	SEQ ID NO: 2392	atttccatgatttctctt	10119	10138	13
SEQ ID NO: 2087	ataatgaataaaaacacat	5608	5627	SEQ ID NO: 2393	atgtcaaacactttgttat	7057	7076	13
SEQ ID NO: 2088	aaacacatctatgacctc	5618	5637	SEQ ID NO: 2394	gatgaagatgacgactttt	12150	12169	13
SEQ ID NO: 2089	gttagaggttcagggttg	5678	5697	SEQ ID NO: 2395	caaatgtctctcaagca	9079	9098	13
SEQ ID NO: 2090	gagtttagcatcgctgctca	5697	5716	SEQ ID NO: 2396	gaggtgactcagagactc	7442	7461	13
SEQ ID NO: 2091	gctgcttcagcacttgac	5732	5751	SEQ ID NO: 2397	gtcagtgaggtctccagc	8888	8907	13
SEQ ID NO: 2092	atttcagcaatgtcttcgc	5782	5801	SEQ ID NO: 2398	cgagcatgggagtgaaat	8620	8639	13
SEQ ID NO: 2093	ttcagcaatgtcttcgc	5783	5802	SEQ ID NO: 2399	acggagcatgggagtgaaa	8618	8638	13
SEQ ID NO: 2094	ttcagcaatgtcttcgc	5784	5803	SEQ ID NO: 2400	acggagcatgggagtgaaa	8618	8637	13
SEQ ID NO: 2095	cagcaatgtcttcgc	5786	5805	SEQ ID NO: 2401	gaaagtgtctcaagctg	12404	12423	13
SEQ ID NO: 2096	gtcttcgc	5792	5811	SEQ ID NO: 2402	atttcaattgggagagaca	6493	6512	13
SEQ ID NO: 2097	gtcttcgc	5793	5812	SEQ ID NO: 2403	ccattcagctctcaagac	12967	12986	13
SEQ ID NO: 2098	attgggaaactgctctc	5851	5870	SEQ ID NO: 2404	cagataaaaaaacctacct	12205	12224	13
SEQ ID NO: 2099	ggagacatactgggcagc	5871	5890	SEQ ID NO: 2405	gtgttttgagactctcc	1080	1099	13
SEQ ID NO: 2100	gttagagcctacagctctg	5906	5925	SEQ ID NO: 2406	cagaattcataatcccaac	8266	8285	13
SEQ ID NO: 2101	gttaggaaagcactcag	5975	5994	SEQ ID NO: 2407	actgcaagattttcagac	13604	13623	13
SEQ ID NO: 2102	agcatcagctcagctctg	5985	6004	SEQ ID NO: 2408	caagagctgttagttgct	13343	13362	13
SEQ ID NO: 2103	tgacacaaagtcagctg	6001	6020	SEQ ID NO: 2409	gcacatcaattgtgtaa	6410	6429	13
SEQ ID NO: 2104	gcagacaggcactctgaaa	6038	6057	SEQ ID NO: 2410	tttcatggcattgtctgc	11602	11621	13
SEQ ID NO: 2105	gaaactcaagccaattt	6053	6072	SEQ ID NO: 2411	aaa tccatcaggttttc	8029	8048	13
SEQ ID NO: 2106	acaattgaatcacagcagga	6076	6095	SEQ ID NO: 2412	tcctttggctgtgtgtgt	9674	9693	13
SEQ ID NO: 2107	gttagctctcaacact	6095	6114	SEQ ID NO: 2413	agtgagttctccagcaag	8591	8610	13
SEQ ID NO: 2108	gttggtggagcttactcg	6124	6143	SEQ ID NO: 2414	ccagaattcataatcccaa	8266	8284	13
SEQ ID NO: 2109	caatttactcagtgagcc	6190	6209	SEQ ID NO: 2415	ggctatgtatgttagagtg	6980	6999	13
SEQ ID NO: 2110	tttagagatgagagatgcc	6227	6246	SEQ ID NO: 2416	ggcctgatgtctattttaa	9169	9188	13
SEQ ID NO: 2111	gagaagccccaagaattta	6249	6268	SEQ ID NO: 2417	aaagccattcagctctc	12962	12981	13
SEQ ID NO: 2112	caattgtgtctttgttaa	6268	6287	SEQ ID NO: 2418	tttaacagctcagatattg	10179	10198	13
SEQ ID NO: 2113	ttgttaagatgataaaa	6278	6297	SEQ ID NO: 2419	ttattgtcgaatcaaaaa	13647	13666	13
SEQ ID NO: 2114	gttaagatgataaaaaa	6280	6299	SEQ ID NO: 2420	tttagaggaatcgcacaa	6350	6369	13
SEQ ID NO: 2115	cttaccatcaactccc	6307	6326	SEQ ID NO: 2421	gggaaaaaacagcgttgaa	9558	9577	13
SEQ ID NO: 2116	tttgagaccttgcaagaa	6329	6348	SEQ ID NO: 2422	ttctctatgggaaaaaa	9558	9577	13
SEQ ID NO: 2117	accttgcaagaattttg	6336	6355	SEQ ID NO: 2423	caaaagaagcccaagaggt	12940	12959	13
SEQ ID NO: 2118	caaatgtcgaattttg	6415	6434	SEQ ID NO: 2424	ccaaagcagattatgtga	11821	11840	13
SEQ ID NO: 2119	cagagcagccctgggaaaa	6443	6462	SEQ ID NO: 2425	ttttagcaccactctctg	13614	13633	13
SEQ ID NO: 2120	ctggggaaaactcccacag	6452	6471	SEQ ID NO: 2426	gtgtcttggtcagccagg	7718	7735	13
SEQ ID NO: 2121	actccacagcagactaat	6461	6480	SEQ ID NO: 2427	attacactctcttgaggt	12861	12880	13
SEQ ID NO: 2122	aattcattcaattggagaa	6489	6508	SEQ ID NO: 2428	ctctctctccatggaatt	10471	10490	13
SEQ ID NO: 2123	tcaattgggagagacaa	6495	6514	SEQ ID NO: 2429	ctggagtgccagttgaa	11800	11819	13
SEQ ID NO: 2124	aggagaaactgactcctc	6526	6545	SEQ ID NO: 2430	agagcttatgggatttctc	11155	11174	13
SEQ ID NO: 2125	actgactgctctcaaaaa	6533	6552	SEQ ID NO: 2431	ttttggcaagctatcacagt	8372	8391	13
SEQ ID NO: 2126	gactgctctcaaaaaaag	6536	6555	SEQ ID NO: 2432	ttttgttgattctcagtc	9687	9706	13
SEQ ID NO: 2127	gacacatatatgataaat	6633	6652	SEQ ID NO: 2433	attggatatacagatctg	1925	1944	13
SEQ ID NO: 2128	aatttgatcagatatataa	6649	6668	SEQ ID NO: 2434	taaaagaatcttcaatt	13807	13826	13
SEQ ID NO: 2129	atgatttaccatgatttga	6675	6694	SEQ ID NO: 2435	taaatgatatcccata	13120	13139	13

SEQ ID NO: 2130	ttgaaaatagctattgct	6689	6708	SEQ ID NO: 2436	agcagcagaaaaaatcaaa	13856	13875	13
SEQ ID NO: 2131	ttgaaatagctattgcta	6690	6709	SEQ ID NO: 2437	ttgacagcagaaaaatcaa	13855	13874	13
SEQ ID NO: 2132	aaatagctattgctaatt	6695	6714	SEQ ID NO: 2438	aaataaaggatcttatt	14076	14095	13
SEQ ID NO: 2133	attattgtagaaacattg	6711	6730	SEQ ID NO: 2439	caataccagaatcataat	8260	8279	13
SEQ ID NO: 2134	aaagctctgtagagcacta	6739	6758	SEQ ID NO: 2440	ttgtagtataacttcctt	12856	12875	13
SEQ ID NO: 2135	aaagctctgtagagcactat	6740	6759	SEQ ID NO: 2441	atagcaaacactaataact	8761	8780	13
SEQ ID NO: 2136	ttgtagagcactatcatat	6745	6764	SEQ ID NO: 2442	atataccaagatgagatcaa	13093	13112	13
SEQ ID NO: 2137	aaatttttagtaaaaaacaat	6769	6788	SEQ ID NO: 2443	attgagattccctcatta	11694	11713	13
SEQ ID NO: 2138	tttagtaaaaaacaatcca	6772	6791	SEQ ID NO: 2444	tggaagtgccagttgaaaa	11802	11821	13
SEQ ID NO: 2139	acattgtttattgaaat	6797	6816	SEQ ID NO: 2445	atttcttaaaagctggatgt	11167	11186	13
SEQ ID NO: 2140	attgatttttaacaaaagtg	6816	6835	SEQ ID NO: 2446	caactgtccagttgtcaat	9863	9882	13
SEQ ID NO: 2141	attttaacaaaagtggaag	6820	6839	SEQ ID NO: 2447	cttcaagcttaaaaaaat	8006	8025	13
SEQ ID NO: 2142	aaatcagaaatccagatata	6880	6899	SEQ ID NO: 2448	ttgatacaaaagccatattt	10080	10099	13
SEQ ID NO: 2143	gaatccagatatacagaasa	6886	6905	SEQ ID NO: 2449	ttttcaactgtgaatttc	9057	9076	13
SEQ ID NO: 2144	taagagacacatcagaaa	6916	6935	SEQ ID NO: 2450	ttcttaaaactctctttaa	9483	9502	13
SEQ ID NO: 2145	atccagcactagctggaa	6942	6961	SEQ ID NO: 2451	ttccaattccctcttgatg	3680	3699	13
SEQ ID NO: 2146	gagcatgtcaaacacttt	7052	7071	SEQ ID NO: 2452	aaagtgcacattttactca	6183	6202	13
SEQ ID NO: 2147	gagcatgtcaaacactttg	7053	7072	SEQ ID NO: 2453	caaatgacatgatgggtct	5326	5345	13
SEQ ID NO: 2148	aaacactttgtttaaatc	7062	7081	SEQ ID NO: 2454	gattatataccatattgttt	13125	13144	13
SEQ ID NO: 2149	tgagaaaataaatgacctt	7103	7122	SEQ ID NO: 2455	gaaggaaaagcgacactca	12021	12040	13
SEQ ID NO: 2150	ttatgaatgagcacacaaca	7152	7171	SEQ ID NO: 2456	ttgtgggggtgatcata	10323	10342	13
SEQ ID NO: 2151	aaagtagcaacaaatcca	7156	7175	SEQ ID NO: 2457	tggaatgagtagcagactt	12148	12167	13
SEQ ID NO: 2152	aaatttgaggagcacttc	7215	7234	SEQ ID NO: 2458	gaataccaactgtcgaactt	10160	10179	13
SEQ ID NO: 2153	acaagtttaagataaaaagat	7256	7275	SEQ ID NO: 2459	acttaaatcagttcttgtt	11326	11345	13
SEQ ID NO: 2154	aaagataaaaagtactttc	7263	7282	SEQ ID NO: 2460	caaaatgaggaaggaactt	2069	2088	13
SEQ ID NO: 2155	gattactttgagaataatg	7272	7291	SEQ ID NO: 2461	taaacattgaattcaact	9061	9080	13
SEQ ID NO: 2156	tgagaaattagtgtgattt	7280	7299	SEQ ID NO: 2462	aaatccgttgaggtagactca	7435	7454	13
SEQ ID NO: 2157	aaattagtgtgatttattg	7284	7303	SEQ ID NO: 2463	caatttgagaatgaattt	10411	10430	13
SEQ ID NO: 2158	tgattttattgtagtgct	7292	7311	SEQ ID NO: 2464	agcatgcctagttttctcca	9945	9964	13
SEQ ID NO: 2159	catgtgaagatgttaacaa	7345	7364	SEQ ID NO: 2465	ttgtagatgaaaccaatga	7414	7433	13
SEQ ID NO: 2160	catgtgaagatgttaacaaa	7346	7365	SEQ ID NO: 2466	ttgttagtgaaaccaatg	7413	7432	13
SEQ ID NO: 2161	attgaaagatgttaacaaat	7347	7366	SEQ ID NO: 2467	attaaagtattgattcaat	10487	10506	13
SEQ ID NO: 2162	tgaaagatgttaacaaatt	7348	7367	SEQ ID NO: 2468	aaattaaagtattgattcaa	10486	10505	13
SEQ ID NO: 2163	gaagatgttaacaaattc	7349	7368	SEQ ID NO: 2469	gaatttaagtattgattcca	10485	10504	13
SEQ ID NO: 2164	aatgttgataaagaattt	7372	7391	SEQ ID NO: 2470	aatccctgaagttgatgt	11478	11498	13
SEQ ID NO: 2165	ttgattaccaacagttgt	7398	7417	SEQ ID NO: 2471	caaatgaaatcccccacaa	8783	8802	13
SEQ ID NO: 2166	caaaatccgttgaggtagct	7433	7452	SEQ ID NO: 2472	agtcctccctcaacagatttg	7964	7983	13
SEQ ID NO: 2167	aaaatccgttgaggtagct	7434	7453	SEQ ID NO: 2473	gaatgaaatcgctgttttt	8630	8649	13
SEQ ID NO: 2168	aggtgactcagagactcaa	7444	7463	SEQ ID NO: 2474	ttgatgatatttggaactct	10723	10742	13
SEQ ID NO: 2169	tgaaatcaggctctcgga	7465	7484	SEQ ID NO: 2475	lccaattctctcttttccac	8401	8420	13
SEQ ID NO: 2170	ttgctgagtatctggaaa	7539	7558	SEQ ID NO: 2476	tttcaagcaaatgacacaac	8532	8551	13
SEQ ID NO: 2171	taagttcagatctttgg	7608	7627	SEQ ID NO: 2477	ccaatgctgaacttttaa	10165	10184	13
SEQ ID NO: 2172	gaaggccaattccggaga	7633	7652	SEQ ID NO: 2478	ttctctttctcatcttca	10205	10224	13
SEQ ID NO: 2173	aatgtatcaaaaggacatt	7676	7695	SEQ ID NO: 2479	aatgaaagtcggatcatt	11013	11032	13
SEQ ID NO: 2174	attcagcaggaaactcaac	7692	7711	SEQ ID NO: 2480	gttgtagagagccccaagat	6246	6265	13
SEQ ID NO: 2175	acctgtctctggtagcaga	7714	7733	SEQ ID NO: 2481	tggaagaaagtgctaggt	9369	9388	13
SEQ ID NO: 2176	ctgtctctggtagcagg	7715	7734	SEQ ID NO: 2482	ctggaactctctagtcagg	8802	8821	13
SEQ ID NO: 2177	gttagcaggatttatagc	7724	7743	SEQ ID NO: 2483	gctaaagggagcagttgacc	10527	10546	13
SEQ ID NO: 2178	ccaggtttatagacacatt	7730	7749	SEQ ID NO: 2484	aaagtcaggtattctctgg	11017	11036	13
SEQ ID NO: 2179	gtttatagacacattgtga	7734	7753	SEQ ID NO: 2485	ggacctgtccattcaaac	13673	13692	13
SEQ ID NO: 2180	actgttcaactacatttct	7745	7764	SEQ ID NO: 2486	ggaaaaaggggattgaagt	10275	10294	13
SEQ ID NO: 2181	ctgattggtgactcttgc	7762	7781	SEQ ID NO: 2487	tgcaagtaagaaaaatcag	14018	14037	13

SEQ ID NO: 2182	atgaagcatgtgttagagc	7839	7858	SEQ ID NO: 2488	gctcatctccttctcat		10200	10219	13
SEQ ID NO: 2183	gaagcatgtgttagagca	7840	7859	SEQ ID NO: 2489	gctcatctccttctca		10199	10218	13
SEQ ID NO: 2184	gggttcactgttctgaaa	7860	7879	SEQ ID NO: 2490	ttaccatagaaggacc		8951	8970	13
SEQ ID NO: 2185	ctcaagaccatcttggagc	7879	7898	SEQ ID NO: 2491	gtccccctaacagattga		7965	7984	13
SEQ ID NO: 2186	ccttgggacatgcctgcg	7889	7908	SEQ ID NO: 2492	ggcaccaggctcgggaag		13970	13989	13
SEQ ID NO: 2187	ttcaggctctcctgaagc	7921	7940	SEQ ID NO: 2493	cttgaaggaaattctgaa		9580	9599	13
SEQ ID NO: 2188	tcagataaactcaaga	7996	8015	SEQ ID NO: 2494	cttcaatgaattcaatgaa		13175	13194	13
SEQ ID NO: 2189	acttcaagaacttaaaaa	8005	8024	SEQ ID NO: 2495	tttaacaaaagggtgaat		6821	6840	13
SEQ ID NO: 2190	atccatccatcaggtttcca	8031	8050	SEQ ID NO: 2496	gggagaagcaaatctggat		9454	9483	13
SEQ ID NO: 2191	gaattaccatcttcaaca	8055	8074	SEQ ID NO: 2497	gttgaagttctccattc		9881	9900	13
SEQ ID NO: 2192	cattctctcttcaaat	8081	8100	SEQ ID NO: 2498	aatccaatttgagaatg		10406	10425	13
SEQ ID NO: 2193	gttgcacgatgtgcaagc	8137	8156	SEQ ID NO: 2499	ctgtgaagaattatcaaa		12924	12943	13
SEQ ID NO: 2194	aatcaccctgcagacatc	8225	8244	SEQ ID NO: 2500	gaagttctcaattttgatt		8514	8533	13
SEQ ID NO: 2195	ggactctcacatccagaa	8312	8331	SEQ ID NO: 2501	ttctctggaaaagggtca		8876	8895	13
SEQ ID NO: 2196	ttcagctctcccatct	8331	8350	SEQ ID NO: 2502	agattctcagatgaggaa		8913	8932	13
SEQ ID NO: 2197	agctatatacagtaattcga	8379	8398	SEQ ID NO: 2503	tcagatggcatgtcgtt		11604	11623	13
SEQ ID NO: 2198	attctgaaattccaactc	8391	8410	SEQ ID NO: 2504	gagataaccgtgcctgaat		11544	11563	13
SEQ ID NO: 2199	ttcacattagatgcaaat	8414	8433	SEQ ID NO: 2505	atttggaaaaaacagaaaa		9730	9749	13
SEQ ID NO: 2200	caaatctgacataggaa	8428	8447	SEQ ID NO: 2506	ttccatcaaaattccttg		9662	9681	13
SEQ ID NO: 2201	gagagctcaaatagaagt	8500	8519	SEQ ID NO: 2507	actttacttccaacitc		13402	13421	13
SEQ ID NO: 2202	agatgccaaatagaagt	8501	8520	SEQ ID NO: 2508	aaactttacttccaactc		13401	13420	13
SEQ ID NO: 2203	ttcaaatlttgatttcaaa	8519	8538	SEQ ID NO: 2509	ttgatcccttlttgaga		11529	11548	13
SEQ ID NO: 2204	caattlttgatttcaagca	8522	8541	SEQ ID NO: 2510	tgctgaattcaaaaagattg		13652	13671	13
SEQ ID NO: 2205	aatgcacaaactccaacc	8541	8560	SEQ ID NO: 2511	gttttcaagggggccatt		12452	12471	13
SEQ ID NO: 2206	agttcttccagcaagtcac	8596	8615	SEQ ID NO: 2512	aggttccatgtgcacact		11380	11399	13
SEQ ID NO: 2207	gttactgcagacaggagca	8608	8627	SEQ ID NO: 2513	gtctcaggagaaactact		13772	13791	13
SEQ ID NO: 2208	tcacaacacagtcgcaagt	8670	8689	SEQ ID NO: 2514	aaactctcaagtcagttga		13414	13433	13
SEQ ID NO: 2209	tcacaactcagcttaccgtga	8743	8762	SEQ ID NO: 2515	ttcattctgtaatatftg		13372	13391	13
SEQ ID NO: 2210	ctggatagcaacactaaat	8757	8776	SEQ ID NO: 2516	attttctgaacttcccag		12694	12713	13
SEQ ID NO: 2211	ctgactgcgcacagat	8821	8840	SEQ ID NO: 2517	ttctgtatgcggaaactcag		12251	12270	13
SEQ ID NO: 2212	agatgagggaacacatgaa	8921	8940	SEQ ID NO: 2518	ttcatgtccctagaaatct		10030	10049	13
SEQ ID NO: 2213	caacttttcaaatctga	9052	9071	SEQ ID NO: 2519	tcaggagataacgtgttga		12610	12629	13
SEQ ID NO: 2214	tttcaaatctgaattcaaa	9059	9078	SEQ ID NO: 2520	ttgatgatctgtccaagaa		7300	7319	13
SEQ ID NO: 2215	gaattcaattcaaacagtcg	9069	9088	SEQ ID NO: 2521	cgacgaagaaaataatttc		13558	13577	13
SEQ ID NO: 2216	cactgtttggagaaaggga	9133	9152	SEQ ID NO: 2522	ttccagaagcagccagtg		12498	12517	13
SEQ ID NO: 2217	actgtttggagaaaggga	9134	9153	SEQ ID NO: 2523	cttcccaagagacagct		2890	2909	13
SEQ ID NO: 2218	aaattctcttttttcag	9213	9232	SEQ ID NO: 2524	ctgatactatgaaaatt		13630	13649	13
SEQ ID NO: 2219	tttttttcagccagccat	9222	9241	SEQ ID NO: 2525	gttgaaggggaagagaaa		13486	13505	13
SEQ ID NO: 2220	tttgaagaagtcgtttcca	9276	9294	SEQ ID NO: 2526	gggaagtcaggtggcaca		10372	10391	13
SEQ ID NO: 2221	cagggaagatagacttct	9304	9323	SEQ ID NO: 2527	aggactcttcaaatctctg		9840	9859	13
SEQ ID NO: 2222	ataagtaacaaaattt	9397	9416	SEQ ID NO: 2528	baatcaggatctgattat		14030	14049	13
SEQ ID NO: 2223	acaacgcagaaactatttga	9427	9446	SEQ ID NO: 2529	ttcattctgaatatattgt		13372	13391	13
SEQ ID NO: 2224	aggaataaaatggagaagca	9455	9474	SEQ ID NO: 2530	tgctgaatttgatctact		11726	11745	13
SEQ ID NO: 2225	agcaaatctgatttctta	9470	9489	SEQ ID NO: 2531	taagttctctgtaactcgt		11711	11730	13
SEQ ID NO: 2226	ttctttaaacaattcttgaa	9494	9513	SEQ ID NO: 2532	ttcaaacgcagcttcagga		13198	13217	13
SEQ ID NO: 2227	tttaacaattctcgaatg	9497	9516	SEQ ID NO: 2533	acttgatttaagtgttaa		9613	9632	13
SEQ ID NO: 2228	acacaataatcacaactc	9526	9545	SEQ ID NO: 2534	ggagacagcatcttgtgt		11203	11222	13
SEQ ID NO: 2229	gaagattttctctatggga	9553	9572	SEQ ID NO: 2535	ttccagaacaaactcttct		3928	3947	13
SEQ ID NO: 2230	gaaaaaaaacagcttggaag	9570	9589	SEQ ID NO: 2536	cttttcaacttcttctt		13013	13032	13
SEQ ID NO: 2231	ttgaaggaaattcttga	9582	9601	SEQ ID NO: 2537	tttgagaatgaatttcaa		10414	10433	13
SEQ ID NO: 2232	gaaggaaattcttga	9583	9602	SEQ ID NO: 2538	gttttggtgataaaattca		11283	11302	13
SEQ ID NO: 2233	agctcagataagaaaaac	9632	9651	SEQ ID NO: 2539	gttgaatgaatacaagct		9797	9816	13

SEQ ID NO: 2234	caaatccttgacagcca	9712	9731	SEQ ID NO: 2540	tgctgcagcagcattga	11680	11699	13
SEQ ID NO: 2235	gatgaacaataaattga	9781	9800	SEQ ID NO: 2541	taacttgcatctgtcat	12754	12773	13
SEQ ID NO: 2236	gaattccgtgatacactgt	9851	9870	SEQ ID NO: 2542	aaacacgaatcacaatt	8930	8949	13
SEQ ID NO: 2237	ttcaggttgcaatgttga	9868	9887	SEQ ID NO: 2543	taaaaagagcttcaggaa	13199	13218	13
SEQ ID NO: 2238	aagtgcttccattcccat	9886	9905	SEQ ID NO: 2544	atgggagatgaagaact	4834	4853	13
SEQ ID NO: 2239	gtcagcactgctactttc	9942	9961	SEQ ID NO: 2545	agaagaagcgcacactgac	11072	11091	13
SEQ ID NO: 2240	ctgccatgggcaatattac	10105	10124	SEQ ID NO: 2546	glaagaaaatcacagacag	6432	6451	13
SEQ ID NO: 2241	tgaaatacaatgctgaact	10159	10178	SEQ ID NO: 2547	agttgaagagacattacca	7216	7235	13
SEQ ID NO: 2242	attgttgctcatctctc	10193	10212	SEQ ID NO: 2548	agaagaacataaactaata	12881	12900	13
SEQ ID NO: 2243	gttgctcatctctcttct	10198	10215	SEQ ID NO: 2549	agaagaatctgcagaaaca	12423	12442	13
SEQ ID NO: 2244	ctgctcattgatgcactgc	10224	10243	SEQ ID NO: 2550	gcagtagactataaagcaga	13920	13939	13
SEQ ID NO: 2245	ccacagcctgctctcag	10297	10316	SEQ ID NO: 2551	ctcaggatctgaagggtgg	8187	8206	13
SEQ ID NO: 2246	atttgctggagggtgactcat	10322	10341	SEQ ID NO: 2552	atgaagtagaccaacaat	7153	7172	13
SEQ ID NO: 2247	atatggaagtgcagcggc	10369	10388	SEQ ID NO: 2553	gccacactccaagcatat	10770	10789	13
SEQ ID NO: 2248	tgaaataccagatcaca	10445	10464	SEQ ID NO: 2554	tttacaattctattcca	13015	13034	13
SEQ ID NO: 2249	aagtcacaaactcactgtct	10455	10474	SEQ ID NO: 2555	agacactagtgattacact	12851	12870	13
SEQ ID NO: 2250	acgtctctctctccatgg	10467	10486	SEQ ID NO: 2556	ctcagtcgaagtcgcccagt	10916	10935	13
SEQ ID NO: 2251	ctctccatggaatttaa	10474	10493	SEQ ID NO: 2557	taatcgagagagatgaag	7140	7159	13
SEQ ID NO: 2252	attcttcaatgctgactc	10504	10523	SEQ ID NO: 2558	gagttggagggtcgggaa	12234	12253	13
SEQ ID NO: 2253	tgaccacaagcttagctt	10540	10559	SEQ ID NO: 2559	aagcgacactcaatataca	12028	12047	13
SEQ ID NO: 2254	ctccactcttactttccc	10565	10584	SEQ ID NO: 2560	ggaactatgctgtagagg	10641	10660	13
SEQ ID NO: 2255	agctgcagggcacttccaa	10702	10721	SEQ ID NO: 2561	ttgggaagaagagcgact	12281	12300	13
SEQ ID NO: 2256	ttcaaaatgtagatatac	10715	10734	SEQ ID NO: 2562	gatatacactaggaggaa	12737	12756	13
SEQ ID NO: 2257	gagaaacatacaagcaaac	10852	10871	SEQ ID NO: 2563	gcttggtttgctcagctc	2459	2478	13
SEQ ID NO: 2258	atggcgaagtgcactgtct	10889	10908	SEQ ID NO: 2564	agaaggtattgaagccat	12952	12971	13
SEQ ID NO: 2259	ggcaaatgctcagctctg	10890	10909	SEQ ID NO: 2565	caagcaggtattaaagcca	12951	12970	13
SEQ ID NO: 2260	ttgctcaggttcacgcaag	10906	10925	SEQ ID NO: 2566	cttgggggaggaggacaa	14058	14077	13
SEQ ID NO: 2261	tgctcagctcagtgcaagt	10907	10926	SEQ ID NO: 2567	acttgggggaggaggacaa	14057	14076	13
SEQ ID NO: 2262	agttcttccatgattccc	10932	10951	SEQ ID NO: 2568	ggaactctgtagaggaaact	12248	12267	13
SEQ ID NO: 2263	tgctaacacataagaaccag	10979	10998	SEQ ID NO: 2569	ctggatgtlaaccaccagca	11178	11197	13
SEQ ID NO: 2264	actaagaaccagaagatca	10986	11005	SEQ ID NO: 2570	tgatcaagaactctgttag	13339	13358	13
SEQ ID NO: 2265	ctaagaaccagagaatcag	10987	11006	SEQ ID NO: 2571	ctgatcaagaactctgttag	13338	13357	13
SEQ ID NO: 2266	cagaagatcagatgga	10995	11014	SEQ ID NO: 2572	tttccagaccaactctctg	13614	13633	13
SEQ ID NO: 2267	aaaaagaagtcgggattc	11010	11029	SEQ ID NO: 2573	gaatttgaagatttgtttt	9272	9291	13
SEQ ID NO: 2268	gattcattctgggtcttct	11024	11043	SEQ ID NO: 2574	gaaaacctatgctctaatc	13158	13177	13
SEQ ID NO: 2269	aagaaaaggcacaacttga	11071	11090	SEQ ID NO: 2575	caaaacctactctgctct	10458	10477	13
SEQ ID NO: 2270	aaggcacactaaaggtctc	11107	11126	SEQ ID NO: 2576	aggacaccaaataaactct	7564	7583	13
SEQ ID NO: 2271	ccagcattggttagagaca	11191	11210	SEQ ID NO: 2577	tgcaacagtaccactcgg	12362	12381	13
SEQ ID NO: 2272	cttggtgacaccaaataac	11231	11250	SEQ ID NO: 2578	gttttaaatgttgtaag	13140	13159	13
SEQ ID NO: 2273	catctcctgtaaaagtttt	11269	11288	SEQ ID NO: 2579	aaaagggtcatggaatgg	8886	8904	13
SEQ ID NO: 2274	gatctaaatcagttctc	11324	11343	SEQ ID NO: 2580	aagatagctagctgtatca	13326	13345	13
SEQ ID NO: 2275	aagaagctgagaactcat	11424	11443	SEQ ID NO: 2581	atgagatcaacacatactt	13102	13121	13
SEQ ID NO: 2276	gttgccctcaactaccaa	11445	11464	SEQ ID NO: 2582	ttggtacagattactcaaa	12633	12652	13
SEQ ID NO: 2277	cttgattccctttttgag	11528	11547	SEQ ID NO: 2583	cttaatttgattttcaag	8520	8539	13
SEQ ID NO: 2278	ttcagcgtctccaaaaggt	11583	11602	SEQ ID NO: 2584	catctatattttctgaa	12685	12704	13
SEQ ID NO: 2279	gtttcagatggcattgct	11600	11619	SEQ ID NO: 2585	agcagattatgttgaaca	11825	11844	13
SEQ ID NO: 2280	aatgcagtagccaacaga	11631	11650	SEQ ID NO: 2586	cttttcagccagccactt	9223	9242	13
SEQ ID NO: 2281	ctgagcagacattgagat	11683	11702	SEQ ID NO: 2587	atctgagtaggaagaactca	12251	12270	13
SEQ ID NO: 2282	ggcagcaccattgagat	11684	11703	SEQ ID NO: 2588	aatctgtagaggaaactca	12250	12269	13
SEQ ID NO: 2283	tgagattccctcattaa	11695	11714	SEQ ID NO: 2589	taactctcataagttcaa	13171	13190	13
SEQ ID NO: 2284	actggagtgccagtttga	11799	11818	SEQ ID NO: 2590	taattgggagagacacagt	6496	6515	13
SEQ ID NO: 2285	caaatggaagacttcag	11996	12015	SEQ ID NO: 2591	ctgagactctcatcttgg	11430	11449	13

SEQ ID NO: 2286	agccagcggttcacgc	12048	12067	SEQ ID NO: 2592	gacccaagtatagttgct	13278	13297	13
SEQ ID NO: 2287	cagcggttcacgcgtcca	12052	12071	SEQ ID NO: 2593	ggacgtgcaccaaagctg	13952	13971	13
SEQ ID NO: 2288	ctccatctgcgtccacaga	12066	12085	SEQ ID NO: 2594	ctgatatacatcacggag	13703	13722	13
SEQ ID NO: 2289	atgaggaaactcagatcaa	12256	12275	SEQ ID NO: 2595	ttgagttgccaccatcat	11659	11678	13
SEQ ID NO: 2290	aggcagctctcgtctgtct	12292	12311	SEQ ID NO: 2596	agcgaagtcttctgacct	3010	3029	13
SEQ ID NO: 2291	gaagacaacgtgcccaa	12319	12338	SEQ ID NO: 2597	ttggagagacaagtttca	6500	6519	13
SEQ ID NO: 2292	tatgattatgcacaaagt	12354	12373	SEQ ID NO: 2598	actttgcactatgttcata	12755	12774	13
SEQ ID NO: 2293	cattaggcaaatgatgat	12467	12486	SEQ ID NO: 2599	atcaacacaatcttcaatg	13107	13126	13
SEQ ID NO: 2294	ttgactcagggaaggccaag	12576	12595	SEQ ID NO: 2600	ctggtagcagagtactcaa	12632	12651	13
SEQ ID NO: 2295	gaacctgggatatacact	12728	12747	SEQ ID NO: 2601	agtgattacacttcccttc	12857	12876	13
SEQ ID NO: 2296	ctcttcaggttaaggaaa	12869	12888	SEQ ID NO: 2602	tttgcacactgctcagga	13516	13535	13
SEQ ID NO: 2297	gccattcagctctcaaga	12966	12985	SEQ ID NO: 2603	ctctccgtctgtaatggc	5794	5813	13
SEQ ID NO: 2298	gtgtctacgtatcttcagg	12993	13012	SEQ ID NO: 2604	ctctgcaccaaagctggcac	13956	13975	13
SEQ ID NO: 2299	agctgaagagatgaatt	13057	13076	SEQ ID NO: 2605	aattttatcaaaacgagct	13192	13211	13
SEQ ID NO: 2300	aatttctctatcttataa	13072	13091	SEQ ID NO: 2606	taaaagaaacttcaatt	13807	13826	13
SEQ ID NO: 2301	tttaaatgttgaagaa	13142	13161	SEQ ID NO: 2607	ttctctctatgggaaaaaa	9558	9577	13
SEQ ID NO: 2302	taattctcataagttcaat	13172	13191	SEQ ID NO: 2608	attgagattccctcatta	11694	11713	13
SEQ ID NO: 2303	tattttgatccaagtata	13271	13290	SEQ ID NO: 2609	tataagcagaagcacatat	13929	13948	13
SEQ ID NO: 2304	gaatattatgaacttga	13303	13322	SEQ ID NO: 2610	caacctaatgattttca	8287	8306	13
SEQ ID NO: 2305	caatttctgcacagaata	13434	13453	SEQ ID NO: 2611	tattctcttttccaattg	13826	13845	13
SEQ ID NO: 2306	gagaagattgcagagcttc	13501	13520	SEQ ID NO: 2612	gaattctcaatttatct	13813	13832	13
SEQ ID NO: 2307	gaagaaataatttctgat	13562	13581	SEQ ID NO: 2613	atcagttcagataaacttc	7991	8010	13
SEQ ID NO: 2308	tgacctgtccattcaaaa	13672	13691	SEQ ID NO: 2614	tttgagaatgaatttcaa	10414	10433	13
SEQ ID NO: 2309	caaaactcacacacattt	13685	13704	SEQ ID NO: 2615	aaattctctgacattgtga	7362	7381	13
SEQ ID NO: 2310	ttttttaaagaattcttc	13803	13822	SEQ ID NO: 2616	gaagttgcagtggcaaaaa	10374	10393	13
SEQ ID NO: 2311	aggatctgagttatttgc	14035	14054	SEQ ID NO: 2617	gcaagggttcacgtttcct	7856	7875	13
SEQ ID NO: 2312	ttgtcaaaactggggggg	14049	14068	SEQ ID NO: 2618	ctcccaggaccttcaaa	9834	9853	13

= Match Number

B = Middle Matching Bases

Table 10. Selected palindromic sequences from human ApoB

	Source	Start Index	End Index		Match	Start Index	End Index	#
SEQ ID NO: 2619	ggccatcgcagaagggaag	517	536	SEQ ID NO: 3948	ctctccgtctgtaatggcc	5803	5822	9
SEQ ID NO: 2620	ggccatctcgagatgtcca	4107	4126	SEQ ID NO: 3949	tggaactctctccatggca	10884	10903	8
SEQ ID NO: 2621	catgtcaaacactgttgta	7064	7083	SEQ ID NO: 3950	taacaaactcttgacatg	7366	7385	8
SEQ ID NO: 2622	ttgtatataactctattg	7076	7095	SEQ ID NO: 3951	caatgaatcatcaatgaaa	8998	9017	8
SEQ ID NO: 2623	ctcgaaaagggtcatgga	8888	8907	SEQ ID NO: 3952	ccagctccatttaccaga	11364	11383	8
SEQ ID NO: 2624	cagctctgtgtcagttcca	10908	10927	SEQ ID NO: 3953	tggaacctgcaccaagctg	13960	13979	8
SEQ ID NO: 2625	ggagggtcccccagctctgc	364	383	SEQ ID NO: 3954	gcagccctgggaaaaatcc	6455	6474	7
SEQ ID NO: 2626	ctgttttgaagactctcca	1089	1108	SEQ ID NO: 3955	tggaagggtagtataacag	10335	10354	7
SEQ ID NO: 2627	agtggtctgaagctgttgcga	1305	1324	SEQ ID NO: 3956	tcgacagctctttctgcaact	13516	13535	7
SEQ ID NO: 2628	ccaaataagaagggaatct	2076	2095	SEQ ID NO: 3957	agattctcttgctctttgg	4008	4027	7
SEQ ID NO: 2629	gaagagaagattgaattt	3628	3647	SEQ ID NO: 3958	aaattctctttcttttcca	9220	9239	7
SEQ ID NO: 2630	agtggtggcaacaccagcca	4238	4257	SEQ ID NO: 3959	tgctagttagggccaacact	10657	10676	7
SEQ ID NO: 2631	aaggctccacaagatcatca	5958	5977	SEQ ID NO: 3960	tgatgatatttggaacctt	10732	10751	7
SEQ ID NO: 2632	ctcagcaggatttatagca	7733	7752	SEQ ID NO: 3961	tgtaagaaccttactgac	7789	7808	7
SEQ ID NO: 2633	gtatctcttggaaccttgaa	10735	10754	SEQ ID NO: 3962	ctacgtcttccgaatcca	7871	7890	7
SEQ ID NO: 2634	gtcaagttgagcaatttct	13431	13450	SEQ ID NO: 3963	agaaaaggccaacctgtgac	11080	11099	7
SEQ ID NO: 2635	atccagatggaagaggga	13488	13507	SEQ ID NO: 3964	ttccaatcttccttggtat	3688	3707	7
SEQ ID NO: 2636	attttttgttcaagaaggt	4551	4570	SEQ ID NO: 3965	acttcagagaaatacaaat	11409	11428	6
SEQ ID NO: 2637	ctcgaaaalgtcagctctgg	212	231	SEQ ID NO: 3966	ccagctcttcctgttaccag	8243	8262	6
SEQ ID NO: 2638	accagagaggtctcttcca	1737	1756	SEQ ID NO: 3967	tgaaagtgtgtctctcgtt	5097	5116	6
SEQ ID NO: 2639	aaagaagttctgaagaagt	1964	1983	SEQ ID NO: 3968	attccatcaacaaatctctt	9689	9688	6
SEQ ID NO: 2640	gctcagagcttaagtctcca	3578	3597	SEQ ID NO: 3969	gtatcttaaatgctagtag	11631	11650	6
SEQ ID NO: 2641	atcaaatgttgaactaattg	6422	6441	SEQ ID NO: 3970	caagaagtcgaagattgat	4561	4580	6
SEQ ID NO: 2642	gaattatcttttaaaacat	7334	7353	SEQ ID NO: 3971	taigtgttaacaaaattctc	11502	11521	6
SEQ ID NO: 2643	agaggcccgctgctgtg	138	157	SEQ ID NO: 3972	gccagaagtgtgagatctcg	3515	3534	6
SEQ ID NO: 2644	ccaactatgagctgagag	279	298	SEQ ID NO: 3973	ctctgagcaacaaattgtt	10317	10336	6
SEQ ID NO: 2645	gctgagagttccagtgagg	290	309	SEQ ID NO: 3974	ctccatggcgaattgtcagc	10893	10912	6
SEQ ID NO: 2646	gaagaaaaccaagaactc	456	475	SEQ ID NO: 3975	gagtcattgaggtttctca	4937	4956	6
SEQ ID NO: 2647	ctactatcatctctgaacca	566	585	SEQ ID NO: 3976	tgctcaataagggagtag	12774	12793	6
SEQ ID NO: 2648	ctactatcatctctgaacat	567	586	SEQ ID NO: 3977	attgtatcaagggaagtag	12773	12792	6
SEQ ID NO: 2649	gagacagaagaagccaagc	623	642	SEQ ID NO: 3978	gctgttggttttgcagcttc	2467	2486	6
SEQ ID NO: 2650	catctactttacagctgaac	679	698	SEQ ID NO: 3979	cttgaaacacaagctcagtg	6008	6027	6
SEQ ID NO: 2651	ctgatcagcagcagccaggt	830	849	SEQ ID NO: 3980	actcgggaagtgcttatcag	5245	5264	6
SEQ ID NO: 2652	acttgagcgtlaagagggaag	862	881	SEQ ID NO: 3981	cttcccaaaagagaccagt	2898	2917	6
SEQ ID NO: 2653	agaggaagcatctgtgcaga	873	892	SEQ ID NO: 3982	ctctggcattactttctct	5929	5948	6
SEQ ID NO: 2654	gaagactctccagggaact	1095	1114	SEQ ID NO: 3983	tggtgaaggagacatttcca	7224	7243	6
SEQ ID NO: 2655	ctctgagcaaaatattccag	1129	1148	SEQ ID NO: 3984	ctgggttactgagctgagag	1169	1188	6
SEQ ID NO: 2656	gtgaagcagtcatactctc	1197	1216	SEQ ID NO: 3985	agagctgcagctctctcat	10024	10043	6
SEQ ID NO: 2657	gtgcacagctgattgagg	1217	1236	SEQ ID NO: 3986	ctctctacagctgtgggcaa	4230	4249	6
SEQ ID NO: 2658	agctgattgaggtgtctcag	1224	1243	SEQ ID NO: 3987	ctggattccacattcagct	11855	11874	6
SEQ ID NO: 2659	gtctccactacatctctcc	1286	1305	SEQ ID NO: 3988	ggaggcgttaagltcagca	7609	7628	6
SEQ ID NO: 2660	gaacagctgtgcatgtccaa	1311	1330	SEQ ID NO: 3989	tggaagagacaagtttcca	6508	6527	6
SEQ ID NO: 2661	gacattgtctaattacttga	1511	1530	SEQ ID NO: 3990	tcagaagacttaagcaatgtc	7240	7259	6
SEQ ID NO: 2662	ttcttctcagacttctct	1746	1765	SEQ ID NO: 3991	aggaagagctccaatttgaa	8506	8525	6
SEQ ID NO: 2663	ccaatatctgaaactcaga	1911	1930	SEQ ID NO: 3992	ctgaaatttcaaatgtg	6493	6512	6
SEQ ID NO: 2664	aaagttagtgaagaagatt	1954	1973	SEQ ID NO: 3993	aaactacacctcatctgctt	2140	2159	6

SEQ ID NO: 2665	aaagtagtgaagaagttc	1955	1974	SEQ ID NO: 3994	gaacctctggcattactt	5924	5943	6
SEQ ID NO: 2666	aaagaagttctgaagaagaat	1964	1983	SEQ ID NO: 3995	attctctggttaactacttt	5490	5509	6
SEQ ID NO: 2667	ttgtgctatccaagaagt	2330	2349	SEQ ID NO: 3996	catcttaggcacagcaaaa	5005	5024	6
SEQ ID NO: 2668	tttgagaagctgtatataa	2389	2408	SEQ ID NO: 3997	tttagccatcggctcaaaa	5708	5727	6
SEQ ID NO: 2669	cagaagggtcctcaagaagt	2569	2588	SEQ ID NO: 3998	attctcttaacaattctgt	9500	9519	6
SEQ ID NO: 2670	gggaagggtcctcaagaagt	2570	2589	SEQ ID NO: 3999	catctctttaaacaattct	9499	9518	6
SEQ ID NO: 2671	gaagggtcctcaagaagtac	2572	2591	SEQ ID NO: 4000	gtcagctctcaggctcttc	7922	7941	6
SEQ ID NO: 2672	caagaalgactttttct	2580	2599	SEQ ID NO: 4001	agaaggatggcatttttg	14008	14027	6
SEQ ID NO: 2673	catggagaatgaccttgaa	2611	2630	SEQ ID NO: 4002	tcagagccaaagtcocatg	7127	7146	6
SEQ ID NO: 2674	ggagccaaaggctgtgagtaa	2687	2706	SEQ ID NO: 4003	ttactccaacgcagcctgc	3058	3077	6
SEQ ID NO: 2675	tcattctctcccaaaagag	2892	2911	SEQ ID NO: 4004	ctctctggggcatctatga	5147	5166	6
SEQ ID NO: 2676	acctatgagctccagagag	3173	3192	SEQ ID NO: 4005	ctctcaagaccacagaggt	12984	13003	6
SEQ ID NO: 2677	gggcaaaacgtctctacaga	3373	3392	SEQ ID NO: 4006	ctctgaagaagaacagtgcc	12325	12344	6
SEQ ID NO: 2678	accttggacattcagaaca	3395	3414	SEQ ID NO: 4007	gtgtcctaaggtctcagggt	5683	5702	6
SEQ ID NO: 2679	atgggcgacctaaagtgtg	3437	3456	SEQ ID NO: 4008	ccacaaatagttttaccat	8949	8968	6
SEQ ID NO: 2680	gatgaaggaagaagtgaat	3626	3645	SEQ ID NO: 4009	attccagctctcccacatc	8338	8357	6
SEQ ID NO: 2681	caatgtatgataccacaaaag	3664	3683	SEQ ID NO: 4010	tttttttgaaalgcacatg	8651	8670	6
SEQ ID NO: 2682	atgatatacaaaaaataga	3668	3687	SEQ ID NO: 4011	tcagtgatgggtctctac	4379	4398	6
SEQ ID NO: 2683	gttcagttcatttggaact	4517	4536	SEQ ID NO: 4012	gtctcaagaaggacttaagc	5312	5331	6
SEQ ID NO: 2684	ttgtttgttcaagaagatc	4552	4571	SEQ ID NO: 4013	gacttcagagaaatacaaa	11408	11427	6
SEQ ID NO: 2685	ttgtttgttcaagaagatca	4553	4572	SEQ ID NO: 4014	tgacttcagagaaatacaaa	11407	11426	6
SEQ ID NO: 2686	ggcaatgggaacactgcgt	5854	5873	SEQ ID NO: 4015	agcgcaggaatcacctcgcca	8227	8246	6
SEQ ID NO: 2687	aacctctgcgactttactt	5925	5944	SEQ ID NO: 4016	aaaggagatgtcaagggtt	10607	10626	6
SEQ ID NO: 2688	catattcttctctcaatga	5934	5953	SEQ ID NO: 4017	tcatttgaagaataaagat	7034	7053	6
SEQ ID NO: 2689	aaagtcatgtgccctgcita	6017	6036	SEQ ID NO: 4018	taagaacacttactgattct	7792	7811	6
SEQ ID NO: 2690	ttccatttttgagacctt	6330	6349	SEQ ID NO: 4019	aaaggacttcaggaatggga	12012	12031	6
SEQ ID NO: 2691	catcaatattgtatcaattt	6421	6440	SEQ ID NO: 4020	aaattaaaaagctctgtatg	6740	6759	6
SEQ ID NO: 2692	aaagatagttatgattta	6673	6692	SEQ ID NO: 4021	taaacccaaaactgtttta	9027	9046	6
SEQ ID NO: 2693	tattgatgaatcatgtgaa	6721	6740	SEQ ID NO: 4022	ctcaagaacttaaaaaata	8015	8034	6
SEQ ID NO: 2694	atgatctacattgtttat	6798	6817	SEQ ID NO: 4023	ataaagaataaagactcat	7388	7407	6
SEQ ID NO: 2695	agagacacatcacagaatat	6927	6946	SEQ ID NO: 4024	tatatattgtcagtgccctc	13390	13409	6
SEQ ID NO: 2696	gacacatcacagaatatga	6930	6949	SEQ ID NO: 4025	cttaaaactcagttctgtc	11335	11354	6
SEQ ID NO: 2697	agcagtcgaacactttgt	7062	7081	SEQ ID NO: 4026	caaaagtcagtgcccttgt	6015	6034	6
SEQ ID NO: 2698	tttgagggaacacaaagg	7523	7542	SEQ ID NO: 4027	cttttgctacacacaaaaa	11238	11257	6
SEQ ID NO: 2699	tttagagggaacacaaagg	7524	7543	SEQ ID NO: 4028	gctttgtgtacacacaaaa	11237	11256	6
SEQ ID NO: 2700	ggagatagactctctgaa	9315	9334	SEQ ID NO: 4029	ctcagaataatcgtttttc	12832	12851	6
SEQ ID NO: 2701	cactgtttctgagctccag	9342	9361	SEQ ID NO: 4030	ctgggaactccaagagtg	12531	12550	6
SEQ ID NO: 2702	cacaatctctgtgtgtgtg	9676	9695	SEQ ID NO: 4031	caattctcaaggaattgtg	10071	10090	6
SEQ ID NO: 2703	ttctgtacatacgtttctc	9861	9880	SEQ ID NO: 4032	ggaaactgtgtactcaggaa	12577	12596	6
SEQ ID NO: 2704	gaattctcaagctttctct	10050	10069	SEQ ID NO: 4033	agagccagtgagagcttc	11052	11071	6
SEQ ID NO: 2705	ttttctatcttcatctgt	10218	10237	SEQ ID NO: 4034	acagctgaaagagatgaaa	13063	13082	6
SEQ ID NO: 2706	ttctaccgtgaataaggagacg	10529	10548	SEQ ID NO: 4035	ctgtcagctgttgagtaga	11769	11788	6
SEQ ID NO: 2707	ttaccgtgaataaggagagct	10530	10549	SEQ ID NO: 4036	actcagcagctttgagtag	11768	11787	6
SEQ ID NO: 2708	agggcctctttttcaccaa	10839	10858	SEQ ID NO: 4037	ttggccaggagtgccct	10955	10984	6
SEQ ID NO: 2709	ttctcatcctgttaaaag	11273	11292	SEQ ID NO: 4038	cttttttccacaacaggagaa	10846	10865	6
SEQ ID NO: 2710	gaaaacaaagcagattat	11824	11843	SEQ ID NO: 4039	ataaaactgtcaagattttc	13608	13627	6
SEQ ID NO: 2711	actcatcattgattttct	12690	12709	SEQ ID NO: 4040	ggaataatcaggatctgagt	14035	14054	6

SEQ ID NO:	2712	taaactaatgatgtgaatc	12898	12917	SEQ ID NO:	4041	gattaccaccagcagttta	13586	13605	1	6
SEQ ID NO:	2713	caaaacgagcttcggaag	13208	13227	SEQ ID NO:	4042	cttcgtgaagaataattca	13268	13287	1	6
SEQ ID NO:	2714	gggaataatgtctcagtg	2374	2393	SEQ ID NO:	4043	aaacttactctgaattcca	10670	10689	3	5
SEQ ID NO:	2715	atttgaatatcaagaag	2406	2427	SEQ ID NO:	4044	cttcagagaataatacaatc	11410	11429	3	5
SEQ ID NO:	2716	atttgaatccaagaagt	2409	2426	SEQ ID NO:	4045	acttcagagaataatacaat	11409	11428	3	5
SEQ ID NO:	2717	atcaacagcgcctctdttg	998	1017	SEQ ID NO:	4046	caagaagtcgaagtgtgat	4561	4580	2	5
SEQ ID NO:	2718	ttgttgaagactctccag	1090	1109	SEQ ID NO:	4047	cttgaagaattaaacaaca	6963	6982	2	5
SEQ ID NO:	2719	ccctctgatgatgtgtgt	1332	1351	SEQ ID NO:	4048	caaaagctctgcaccagg	13969	13988	2	5
SEQ ID NO:	2720	tgagcaagtgaagaacttt	1876	1895	SEQ ID NO:	4049	aaagccatctcagctctcca	12971	12990	2	5
SEQ ID NO:	2721	atttgaatccaaagaagt	2409	2428	SEQ ID NO:	4050	actttctaaacttgaaat	9063	9082	2	5
SEQ ID NO:	2722	atccaaagaagctccggaa	2416	2435	SEQ ID NO:	4051	ttccggggaaacctgggat	12729	12748	2	5
SEQ ID NO:	2723	agagcctactctcgactct	2438	2457	SEQ ID NO:	4052	agatgtgtatcgttagcctct	11929	11948	2	5
SEQ ID NO:	2724	aatgccttgaactccoca	2618	2637	SEQ ID NO:	4053	gggaactacataatttact	7020	7039	2	5
SEQ ID NO:	2725	gaagtcacaaactccgatt	3305	3324	SEQ ID NO:	4054	aaactctcaattttcttc	13823	13842	2	5
SEQ ID NO:	2726	tgcaagcagaagccagag	3504	3523	SEQ ID NO:	4055	cttcaggctccacgtgca	11384	11403	2	5
SEQ ID NO:	2727	gaagcagaagatgaattg	3629	3648	SEQ ID NO:	4056	caaaacctctgtctctctc	10467	10486	2	5
SEQ ID NO:	2728	gtgclaaaggccacatgag	4605	4624	SEQ ID NO:	4057	tcatatgaagaatcagcat	12684	12683	2	5
SEQ ID NO:	2729	ttccctacactccactctg	4745	4764	SEQ ID NO:	4058	cagattctcagatgaggga	8920	8939	2	5
SEQ ID NO:	2730	atttcagctctgacaaagt	5435	5454	SEQ ID NO:	4059	actttcttaacttgaaat	9063	9082	2	5
SEQ ID NO:	2731	aggagcctacaaaataat	5602	5621	SEQ ID NO:	4060	atattgtttgaaacagtcct	11838	11857	2	5
SEQ ID NO:	2732	aaagctgaagcacatcaat	6405	6428	SEQ ID NO:	4061	atttgttgcctactctctt	10202	10221	2	5
SEQ ID NO:	2733	ctgctctgaacaacgagaa	9426	9445	SEQ ID NO:	4062	ttctgtaccaccacagag	13582	13601	2	5
SEQ ID NO:	2734	tgaaaggaaattctgaaaa	9590	9609	SEQ ID NO:	4063	tttaaaagaatactctcaa	13813	13832	2	5
SEQ ID NO:	2735	gaagtaaaagaaaattttg	10751	10770	SEQ ID NO:	4064	caaaacctctgtctctctc	10467	10486	2	5
SEQ ID NO:	2736	tgaaagaagatggcaaatit	11992	12011	SEQ ID NO:	4065	aaatgtcagctctgtgtca	10902	10921	2	5
SEQ ID NO:	2737	aggatctgagtatttttgc	14043	14062	SEQ ID NO:	4066	gcaagctagccagcttctct	10928	10947	2	5
SEQ ID NO:	2738	tgccctctctggtgtctg	26	45	SEQ ID NO:	4067	cagccattgacatgagcac	5748	5767	1	5
SEQ ID NO:	2739	ggcgtctgcctgctgctg	154	173	SEQ ID NO:	4068	cagctccacagactccgcc	3070	3089	1	5
SEQ ID NO:	2740	ctgctgctgctgctgctg	162	181	SEQ ID NO:	4069	agcagaaggtgcgaagcag	3232	3251	1	5
SEQ ID NO:	2741	ctgctggtggcgccgcagg	178	197	SEQ ID NO:	4070	ctcgtgattccacatgcagc	11854	11873	1	5
SEQ ID NO:	2742	gaaggagaaactcgtgaaaa	201	220	SEQ ID NO:	4071	ttttcttactacatctt	2592	2611	1	5
SEQ ID NO:	2743	ctggaaaatgtcagccctg	212	231	SEQ ID NO:	4072	ccagactctccacatccagc	3923	3942	1	5
SEQ ID NO:	2744	ggggctcctcgtgactgct	304	323	SEQ ID NO:	4073	agcagctcgtgatttccca	9953	9972	1	5
SEQ ID NO:	2745	ggagctcctcgtgactgctg	305	324	SEQ ID NO:	4074	cagactcgtgatttctcc	9952	9971	1	5
SEQ ID NO:	2746	gggagctcgtgactcaaga	313	332	SEQ ID NO:	4075	ttctctacacactggaacca	2050	2069	1	5
SEQ ID NO:	2747	ctgctgattcaagaatgtc	318	337	SEQ ID NO:	4076	gcacacactctgacatgcag	11087	11106	1	5
SEQ ID NO:	2748	gcaccaggatcaactctg	334	353	SEQ ID NO:	4077	gcaggctgaactgtgtgcca	2725	2744	1	5
SEQ ID NO:	2749	ccaccaggatcaactctg	335	354	SEQ ID NO:	4078	gcaggctgaactgtgtgcca	2724	2743	1	5
SEQ ID NO:	2750	tgcaaggttgagctgtgagg	350	369	SEQ ID NO:	4079	ctccacactctgactgtgca	4752	4771	1	5
SEQ ID NO:	2751	caaggttgagctgtgaggtt	352	371	SEQ ID NO:	4080	aaacctctacatgaagcttg	13763	13782	1	5
SEQ ID NO:	2752	ctctgcagcttactctgga	377	396	SEQ ID NO:	4081	tcaggagactcttcaagag	13219	13238	1	5
SEQ ID NO:	2753	cagcttactctgaagacc	382	401	SEQ ID NO:	4082	ggtcttgagtaaatgtctg	4985	5004	1	5
SEQ ID NO:	2754	cttactctgaagaccagc	384	403	SEQ ID NO:	4083	cttgagcgtctgaagagagc	863	882	1	5
SEQ ID NO:	2755	ctactctgaagaccagcca	387	406	SEQ ID NO:	4084	tgcatggcattatgatga	3612	3631	1	5
SEQ ID NO:	2756	aaaaaccaagaactctgag	460	479	SEQ ID NO:	4085	ctcaacctaatgatatttcc	8294	8313	1	5
SEQ ID NO:	2757	tgaaactctgaggaatttgc	468	487	SEQ ID NO:	4086	gcaagctatacatgattct	8385	8404	1	5
SEQ ID NO:	2758	ctcagaggaattgtctgac	473	492	SEQ ID NO:	4087	ctcaggggatactccacaga	2534	2553	1	5

SEQ ID NO: 2759	tttgcagccagatgcca	482	501	SEQ ID NO: 4088	gggaagtgctagtgcaaa	10380	10399	1	5
SEQ ID NO: 2760	caagaggggacatcttct	586	605	SEQ ID NO: 4089	gaataaagtcgtcttg	7043	7062	1	5
SEQ ID NO: 2761	tcacttaccgtcaagagc	682	701	SEQ ID NO: 4090	cgctacacatcatgtag	4368	4387	1	5
SEQ ID NO: 2762	ttaccgtcaagacgaggaa	686	705	SEQ ID NO: 4091	tccttgacatgtagataa	7374	7393	1	5
SEQ ID NO: 2763	cacgtgacgtcaagaggaa	861	880	SEQ ID NO: 4092	ttccagaagacgcacgtg	12506	12525	1	5
SEQ ID NO: 2764	gggaagcattggtcagaag	875	894	SEQ ID NO: 4093	cttcacacatctaatct	9996	10015	1	5
SEQ ID NO: 2765	caaggagcaaacctcttc	901	920	SEQ ID NO: 4094	gaagtagtagtgcactgt	6843	6862	1	5
SEQ ID NO: 2766	acagactttgaaacttgaa	967	986	SEQ ID NO: 4095	ttcaattctcaatgtct	10508	10527	1	5
SEQ ID NO: 2767	gatgaagcagctacatct	1195	1214	SEQ ID NO: 4096	gagatttgaggatcatca	7984	8003	1	5
SEQ ID NO: 2768	ggcagtcacatctctctg	1201	1220	SEQ ID NO: 4097	caaggagaacactgactgt	6532	6551	1	5
SEQ ID NO: 2769	ccagcccatcactttaca	1239	1258	SEQ ID NO: 4098	gttagtctctgtgtgtgtg	5102	5121	1	5
SEQ ID NO: 2770	ctccatccacatctccag	1288	1307	SEQ ID NO: 4099	ctggagcttagtaagtggag	8717	8736	1	5
SEQ ID NO: 2771	catgccaacccccctctga	1322	1341	SEQ ID NO: 4100	tcagtgagggaacacatg	8927	8946	1	5
SEQ ID NO: 2772	gagagatcttcaacatggc	1398	1417	SEQ ID NO: 4101	gccaccctggaaactctct	10877	10896	1	5
SEQ ID NO: 2773	tcacatgtgcgagggatca	1407	1426	SEQ ID NO: 4102	gtatccaccctctcatgga	2973	2992	1	5
SEQ ID NO: 2774	ccacctgtatgtcgtgag	1437	1456	SEQ ID NO: 4103	ctcagggtatgcaagggtg	8195	8214	1	5
SEQ ID NO: 2775	gtcaacaactatcataaga	1463	1482	SEQ ID NO: 4104	cttgtagttaaattgtgac	4987	5006	1	5
SEQ ID NO: 2776	tgacattgtctaataact	1509	1528	SEQ ID NO: 4105	agatatactgaagaatcca	12807	12826	1	5
SEQ ID NO: 2777	ggacattgcaactatcgtg	1510	1529	SEQ ID NO: 4106	gagtatattctgcgaagtcc	12806	12825	1	5
SEQ ID NO: 2778	ttctgctgggtcatgtgaaa	1581	1600	SEQ ID NO: 4107	tttccatgccaaggaggaa	6522	6541	1	5
SEQ ID NO: 2779	ccagaactcaagctctcaa	1628	1647	SEQ ID NO: 4108	tgaagltgactctctcgg	5096	5115	1	5
SEQ ID NO: 2780	gactctcaatctcgaatg	1638	1657	SEQ ID NO: 4109	cattctgattgtgtgact	7765	7784	1	5
SEQ ID NO: 2781	tgagcaagtgaagaacttt	1876	1895	SEQ ID NO: 4110	aaagtgccatttcaacca	6191	6210	1	5
SEQ ID NO: 2782	gccaagtgaaagacttgt	1878	1897	SEQ ID NO: 4111	caaaagtcagtgccctgt	6015	6034	1	5
SEQ ID NO: 2783	ctgaagaatctcaact	1972	1991	SEQ ID NO: 4112	agttccataatgtgttcaga	12819	12838	1	5
SEQ ID NO: 2784	gactgtatgtgacttcagaa	1994	2013	SEQ ID NO: 4113	ttctgaatatattgtcagt	13384	13403	1	5
SEQ ID NO: 2785	acttgaccgcagctcagcc	2059	2078	SEQ ID NO: 4114	ggctcaccctgagagaagtt	12399	12418	1	5
SEQ ID NO: 2786	tcacaataactactctct	2104	2123	SEQ ID NO: 4115	aggaagatatgaagatgga	4720	4739	1	5
SEQ ID NO: 2787	actaactcactgctcttg	2141	2160	SEQ ID NO: 4116	caaatgttgaggaggtgat	10327	10346	1	5
SEQ ID NO: 2788	ttgattgtctctcagctga	2157	2176	SEQ ID NO: 4117	tcagtataagtcacaccaa	9400	9419	1	5
SEQ ID NO: 2789	ttggaagctctttttggga	2219	2238	SEQ ID NO: 4118	ccogattcacgccttccaa	11585	11604	1	5
SEQ ID NO: 2790	ggaagctcttttgggaag	2221	2240	SEQ ID NO: 4119	cttcagaagctcaactct	7937	7956	1	5
SEQ ID NO: 2791	ttttccagacagctgcca	2246	2265	SEQ ID NO: 4120	gacactctcagaagcaaaa	4884	4903	1	5
SEQ ID NO: 2792	gagacagltgtaacaaagct	2254	2273	SEQ ID NO: 4121	agctgtgtttttccagctt	2466	2485	1	5
SEQ ID NO: 2793	ctttggctataccaaagat	2329	2348	SEQ ID NO: 4122	atctctgtctaggaaagat	5976	5995	1	5
SEQ ID NO: 2794	caaaagatataaacatgag	2341	2360	SEQ ID NO: 4123	ctccaaggataacgtgttg	12617	12636	1	5
SEQ ID NO: 2795	gatalgtgtaaatggaataa	2363	2382	SEQ ID NO: 4124	tatctataatttatatc	13087	13106	1	5
SEQ ID NO: 2796	ggaaatgctcagtgctgtg	2375	2394	SEQ ID NO: 4125	caacactactgaaatcc	10669	10688	1	5
SEQ ID NO: 2797	ttgaaatccaaagaagtc	2410	2429	SEQ ID NO: 4126	gacttcagagaataacaaa	11408	11427	1	5
SEQ ID NO: 2798	gatcccccagatgattgga	2542	2561	SEQ ID NO: 4127	tccaatttccctgtgtgact	3689	3708	1	5
SEQ ID NO: 2799	agatgattggagaggatca	2549	2568	SEQ ID NO: 4128	ggaccacacaaacagctg	5371	5390	1	5
SEQ ID NO: 2800	gaaatgactttttttcca	2583	2602	SEQ ID NO: 4129	gaagctccgattcatct	11023	11042	1	5
SEQ ID NO: 2801	gaaatccccacgtggagctg	2627	2646	SEQ ID NO: 4130	cagctcaacacgtcaggttc	11869	11888	1	5
SEQ ID NO: 2802	ataatctcatctgagatca	2660	2679	SEQ ID NO: 4131	tgacttcagtgacgaatat	11974	11993	1	5
SEQ ID NO: 2803	gtcatgtctcccgagacca	2675	2694	SEQ ID NO: 4132	ggccccgttttaacatgac	5817	5836	1	5
SEQ ID NO: 2804	gtgaaatttatcatctct	2881	2900	SEQ ID NO: 4133	aggaagctttaagttaagc	7608	7627	1	5
SEQ ID NO: 2805	attctctcccaagagagac	2894	2913	SEQ ID NO: 4134	gtctctctcatgtaaat	10478	10497	1	5

SEQ ID NO: 2806	ctcattgagaacaggcagt	2984	3003	SEQ ID NO: 4135	actgactgacgcgttgg	11764	11783	1	5
SEQ ID NO: 2807	tgagcagtatctctgac	3150	3169	SEQ ID NO: 4136	ctgagagaagtgcttcaa	12407	12426	1	5
SEQ ID NO: 2808	acccctgctcagctgaagtc	3293	3312	SEQ ID NO: 4137	ggacggctactgtccaggt	12792	12811	1	5
SEQ ID NO: 2809	ccagtgaaagtcacaattcc	3300	3319	SEQ ID NO: 4138	ggagagcagaggttactcg	9156	9175	1	5
SEQ ID NO: 2810	ccatcagagaacagaaaat	3402	3421	SEQ ID NO: 4139	atttccctaagcgttgaggt	11175	11194	1	5
SEQ ID NO: 2811	gaaaaatcaagggtgttat	3471	3490	SEQ ID NO: 4140	ataaacctgcaagatttctc	13608	13627	1	5
SEQ ID NO: 2812	aaatcaagggtgttatttc	3474	3493	SEQ ID NO: 4141	gaaacaatgcattagattt	9753	9772	1	5
SEQ ID NO: 2813	tgccattatgtagaagaga	3617	3636	SEQ ID NO: 4142	ctcccggttataatgccca	11789	11808	1	5
SEQ ID NO: 2814	aaagagaagtgtaatttga	3630	3649	SEQ ID NO: 4143	caaaaacctactgctctct	10466	10485	1	5
SEQ ID NO: 2815	aaatgacttccaatttccc	3681	3700	SEQ ID NO: 4144	gggaacacataaatttcatt	7021	7040	1	5
SEQ ID NO: 2816	atgacttccaatttccctcg	3683	3702	SEQ ID NO: 4145	caggctgattacagatcat	4925	4944	1	5
SEQ ID NO: 2817	acttccaatttccctgttg	3686	3705	SEQ ID NO: 4146	ccacgcaaaaataatggaggt	10388	10387	1	5
SEQ ID NO: 2818	agtgccaatgagctcattg	3811	3830	SEQ ID NO: 4147	ccatcagttcagataaaact	7997	8016	1	5
SEQ ID NO: 2819	ttgcaagaccacctcaat	3868	3887	SEQ ID NO: 4148	attgacctgcctcatcaaa	13679	13698	1	5
SEQ ID NO: 2820	gaaggagtgcaaacctccag	3992	3911	SEQ ID NO: 4149	ctgtgaattgtcatctcttc	11736	11755	1	5
SEQ ID NO: 2821	acttccatccatccagaaaa	3927	3946	SEQ ID NO: 4150	tttaacaanaagtggaagt	6829	6848	1	5
SEQ ID NO: 2822	ctctcttfaaaaagcagtg	3947	3966	SEQ ID NO: 4151	catcactgccaaggagagag	8494	8513	1	5
SEQ ID NO: 2823	aaaagcagtgatgcccgttca	3956	3975	SEQ ID NO: 4152	tgactcactcattgatttt	12688	12707	1	5
SEQ ID NO: 2824	tcccttgccttttggttg	4011	4030	SEQ ID NO: 4153	ccacaacaacatgaagggaa	9284	9283	1	5
SEQ ID NO: 2825	caagtcctgggattccat	4087	4106	SEQ ID NO: 4154	atggggaaaaaaacgctgtg	9574	9593	1	5
SEQ ID NO: 2826	aagtcctactttttccat	4125	4144	SEQ ID NO: 4155	atgggggaagtataagaactt	4842	4861	1	5
SEQ ID NO: 2827	tgccctctcgggtgttgtt	4167	4186	SEQ ID NO: 4156	agaaaaaacaacacaggcca	9651	9670	1	5
SEQ ID NO: 2828	accagcacagaccatttca	4250	4269	SEQ ID NO: 4157	tgagtgtagtctcttggt	5097	5116	1	5
SEQ ID NO: 2829	ccagcacagaccatttcag	4251	4270	SEQ ID NO: 4158	ctgaataacaaatgctctgg	5519	5538	1	5
SEQ ID NO: 2830	actatcatgtgatgggtct	4375	4394	SEQ ID NO: 4159	agacacctgtatttattagt	7956	7975	1	5
SEQ ID NO: 2831	accacagatgctgcttca	4504	4523	SEQ ID NO: 4160	tgaggctgactctgttggt	4290	4309	1	5
SEQ ID NO: 2832	ccacagatgctgcttcag	4505	4524	SEQ ID NO: 4161	ctgagcaacaatttgggtg	10319	10338	1	5
SEQ ID NO: 2833	ttggactcacaagaagaaa	4528	4547	SEQ ID NO: 4162	ttctctcatgatatacaaa	5941	5960	1	5
SEQ ID NO: 2834	caaagaagctcaagattga	4560	4579	SEQ ID NO: 4163	ccaaggataaagctgtttga	12618	12637	1	5
SEQ ID NO: 2835	atgagaacacagagctgac	4806	4825	SEQ ID NO: 4164	gtcagatatgtgtgcact	10195	10214	1	5
SEQ ID NO: 2836	ttaaaatctgacaccaatg	4826	4845	SEQ ID NO: 4165	cattcattgagaagtgtaa	7350	7369	1	5
SEQ ID NO: 2837	gaagtataagaacttgcgc	4846	4865	SEQ ID NO: 4166	ggcaaatitgaaggacttc	12002	12021	1	5
SEQ ID NO: 2838	aagttataagaacttgcga	4847	4866	SEQ ID NO: 4167	ggccaatitgaaggacttc	12001	12020	1	5
SEQ ID NO: 2839	tcttcagcctgtctcttg	4949	4968	SEQ ID NO: 4168	cagaatccagatatacaaga	6892	6911	1	5
SEQ ID NO: 2840	ctggatcactacaatttcca	4965	4984	SEQ ID NO: 4169	gggttcttccagagaccag	11041	11060	1	5
SEQ ID NO: 2841	aaattaatgtgtgtgtctca	5022	5041	SEQ ID NO: 4170	tgagaagccccaagaattt	6256	6275	1	5
SEQ ID NO: 2842	atgccaacgaccaacttga	5081	5100	SEQ ID NO: 4171	tcaaatctctgtatataact	9856	9875	1	5
SEQ ID NO: 2843	ctgggaagtgcttatcagg	5246	5265	SEQ ID NO: 4172	ccctgacctcacataccag	8318	8337	1	5
SEQ ID NO: 2844	gcaaaaacattttcaactt	5286	5305	SEQ ID NO: 4173	aagtaaaagaagaaatttgc	10752	10771	1	5
SEQ ID NO: 2845	aaaaaacattttcaacttca	5288	5307	SEQ ID NO: 4174	tgaaataaagaagaaattt	10750	10769	1	5
SEQ ID NO: 2846	tcagccaagaaggacttaa	5310	5329	SEQ ID NO: 4175	taaggacttccattctga	13371	13390	1	5
SEQ ID NO: 2847	tcaatgacatgatggcct	5333	5352	SEQ ID NO: 4176	agcccatcaatacttga	6213	6232	1	5
SEQ ID NO: 2848	cacacaacacagctgtga	5375	5394	SEQ ID NO: 4177	gtttcaactgcctttgtg	11227	11246	1	5
SEQ ID NO: 2849	cttcaaaaactgtacaaca	5417	5436	SEQ ID NO: 4178	gttttctatttccaaga	12843	12862	1	5
SEQ ID NO: 2850	caagtttataagcaaac	5449	5468	SEQ ID NO: 4179	agttattttgttaaaactgt	14051	14070	1	5
SEQ ID NO: 2851	tgtaactactttaaacag	5496	5515	SEQ ID NO: 4180	ctgttttatagagaacca	7520	7539	1	5
SEQ ID NO: 2852	taacagtgacctgaataata	5510	5529	SEQ ID NO: 4181	gtatagcaaatcttgtt	5898	5917	1	5

SEQ ID NO: 2853	gggaacatcggtgagac	5552	5571	SEQ ID NO: 4182	gtctctccatgattccc	10941	10960	1	5
SEQ ID NO: 2854	aaacatctatgccccttc	5628	5647	SEQ ID NO: 4183	ggagacgacatcttggtt	11212	11231	1	5
SEQ ID NO: 2855	cagcaagctataaagcag	5660	5679	SEQ ID NO: 4184	ctgtctaaagaccttactga	7788	7807	1	5
SEQ ID NO: 2856	gcagacatgttgtaag	5675	5694	SEQ ID NO: 4185	ctcttaacagcatgactgc	11754	11773	1	5
SEQ ID NO: 2857	ctggggagaacatctg	5674	5693	SEQ ID NO: 4186	ccaggttttccacacacaga	8046	8065	1	5
SEQ ID NO: 2858	ttctctcatgattacaaag	5942	5961	SEQ ID NO: 4187	cttttccaccaacggagaa	10846	10865	1	5
SEQ ID NO: 2859	ctgagcagacagcaccctg	6042	6061	SEQ ID NO: 4188	caggaggccttaagttacg	7607	7626	1	5
SEQ ID NO: 2860	caatttaacacacgaat	6074	6093	SEQ ID NO: 4189	attctctcttcttaacaattg	8090	8109	1	5
SEQ ID NO: 2861	ggacgaactctggctgac	6148	6167	SEQ ID NO: 4190	gtcagccaggtcttcttcca	10932	10951	1	5
SEQ ID NO: 2862	cttttactagtgcagccca	6200	6219	SEQ ID NO: 4191	gggtctaaagctatgaaag	7835	7854	1	5
SEQ ID NO: 2863	ctatgctgcttttagagat	6225	6244	SEQ ID NO: 4192	atctctataagttcaatga	13182	13201	1	5
SEQ ID NO: 2864	aaacccaagctgtttcac	6303	6322	SEQ ID NO: 4193	gagtgtaaatgctgttttt	8638	8657	1	5
SEQ ID NO: 2865	aggaatcgacaaacacatta	6365	6384	SEQ ID NO: 4194	aatgattttcaagttctct	8302	8321	1	5
SEQ ID NO: 2866	agttgtactggaaacagct	6384	6403	SEQ ID NO: 4195	acgttagctcttaagacta	11936	11955	1	5
SEQ ID NO: 2867	ggaaacgtacagagaaag	6394	6413	SEQ ID NO: 4196	cttttcaattcatctttcc	13022	13041	1	5
SEQ ID NO: 2868	gaaacgtacagagaaagc	6395	6414	SEQ ID NO: 4197	gtctctcttccacatttc	10080	10099	1	5
SEQ ID NO: 2869	aaagctgaagcacaatcaat	6409	6428	SEQ ID NO: 4198	attgattgttagagctgttt	6992	7011	1	5
SEQ ID NO: 2870	aaagctgaagcacaatcaat	6410	6429	SEQ ID NO: 4199	attgattgttagagctgttt	6991	7010	1	5
SEQ ID NO: 2871	gaagcacaatcaatgaa	6414	6433	SEQ ID NO: 4200	taaccttaacattgattttca	8295	8314	1	5
SEQ ID NO: 2872	atcaatattgatcaatttg	6422	6441	SEQ ID NO: 4201	caaaagccatcactgatgat	1668	1687	1	5
SEQ ID NO: 2873	aaatgattatctgaattcca	6484	6503	SEQ ID NO: 4202	gaaatcatgtgaaataaag	6727	6746	1	5
SEQ ID NO: 2874	gattatctgaattcatcca	6488	6507	SEQ ID NO: 4203	ggaagtgcctgagaaatac	7102	7121	1	5
SEQ ID NO: 2875	aatgggagagacaaattt	6506	6525	SEQ ID NO: 4204	aaacattctcttaacaatt	9498	9517	1	5
SEQ ID NO: 2876	aaaatagctatgtcctaata	6701	6720	SEQ ID NO: 4205	attgaaataattgattttt	6814	6833	1	5
SEQ ID NO: 2877	aaaatttaaaagctttgat	6739	6758	SEQ ID NO: 4206	atcatatcogtgaatttt	6765	6784	1	5
SEQ ID NO: 2878	tgaaaatattgatttttaa	6816	6835	SEQ ID NO: 4207	taattctcataagttcaa	13179	13198	1	5
SEQ ID NO: 2879	agacatccagcacctagct	6946	6965	SEQ ID NO: 4208	agcttggttttgccagctt	2466	2485	1	5
SEQ ID NO: 2880	caatttcatttgaagaat	7029	7048	SEQ ID NO: 4209	attctctcttcaacaattg	8090	8109	1	5
SEQ ID NO: 2881	agtttttaalgataaatt	7182	7201	SEQ ID NO: 4210	aatgtgtgaaagaaaacct	13155	13174	1	5
SEQ ID NO: 2882	cagaagctaaagcattgccc	7241	7260	SEQ ID NO: 4211	ggcaaggcccagaattctg	12553	12572	1	5
SEQ ID NO: 2883	gagataaaagattacttt	7270	7289	SEQ ID NO: 4212	aaagaaaacctatgcttta	13163	13182	1	5
SEQ ID NO: 2884	aaagattactttgagaatt	7277	7296	SEQ ID NO: 4213	attctcttaaacattcttt	9489	9508	1	5
SEQ ID NO: 2885	gagaatattgttgatttaa	7289	7308	SEQ ID NO: 4214	aaagcattatgactcttc	12670	12689	1	5
SEQ ID NO: 2886	atttatgtagatgcgtc	7303	7322	SEQ ID NO: 4215	gcacatgttataagaagaat	7379	7398	1	5
SEQ ID NO: 2887	gaattctctttaaacaat	7334	7353	SEQ ID NO: 4216	atgtatcaaatggaacttc	7685	7704	1	5
SEQ ID NO: 2888	ttaccacagttgttagat	7411	7430	SEQ ID NO: 4217	atctggaaacctggaatga	10739	10758	1	5
SEQ ID NO: 2889	ttgcagtgatctggaaag	7548	7567	SEQ ID NO: 4218	cttttccactatgagtgcaa	8420	8439	1	5
SEQ ID NO: 2890	cattcagcaggaaacttcaa	7699	7718	SEQ ID NO: 4219	tgagggaactcaggagatt	12009	12028	1	5
SEQ ID NO: 2891	ccacctgattttatgtccc	7958	7977	SEQ ID NO: 4220	ggacactcaagataactgtg	12614	12633	1	5
SEQ ID NO: 2892	ggattccatcagctcagat	7992	8011	SEQ ID NO: 4221	atcttcaatgattatattc	13124	13143	1	5
SEQ ID NO: 2893	ttgagaatgaaagttaa	8112	8131	SEQ ID NO: 4222	ttatgattatgtcaacaa	12360	12379	1	5
SEQ ID NO: 2894	tgacacagtgagctcagtt	8156	8175	SEQ ID NO: 4223	actggactctctgactcag	8809	8828	1	5
SEQ ID NO: 2895	atcaaatctctctcttttc	8407	8426	SEQ ID NO: 4224	gaaaatgaagtcogatt	11017	11036	1	5
SEQ ID NO: 2896	tttttatttcaacgcaaa	8532	8551	SEQ ID NO: 4225	ttgcaagtttaagaaat	14023	14042	1	5
SEQ ID NO: 2897	tttttatttcaacgcaaat	8533	8552	SEQ ID NO: 4226	atttgatttaagtgttaa	9622	9641	1	5
SEQ ID NO: 2898	gtatttcaagcaaatgca	8536	8555	SEQ ID NO: 4227	tgcaagtttaagaaatca	14025	14044	1	5
SEQ ID NO: 2899	ttctctgttttttggaaatg	8645	8664	SEQ ID NO: 4228	atttgtagagagacagcat	11203	11222	1	5

SEQ ID NO: 2900	gcctgtttttggaatgc	8646	8665	SEQ ID NO: 4229	gcattgttgaggagacatg	11202	11221	5
SEQ ID NO: 2901	aaaaaaatcatctggagct	8706	8725	SEQ ID NO: 4230	agctagagggtcttttt	10833	10851	5
SEQ ID NO: 2902	actggagcttagtaattga	8716	8735	SEQ ID NO: 4231	tccactcacatctccagt	1289	1308	5
SEQ ID NO: 2903	cttctcgaaagggtctatg	8886	8905	SEQ ID NO: 4232	catgaaacctcatgtgag	13759	13778	5
SEQ ID NO: 2904	ggaaaagggtctatgaaat	8891	8910	SEQ ID NO: 4233	atttgaaagctgtttcc	9282	9301	5
SEQ ID NO: 2905	ggggcgtcccgacattctc	8910	8929	SEQ ID NO: 4234	ggaaacattatggaggccc	9440	9459	5
SEQ ID NO: 2906	tctcatgagtgagggaacac	8924	8943	SEQ ID NO: 4235	gttcttcaagcttgagaa	12416	12435	5
SEQ ID NO: 2907	atgtgagggaacacatgaat	8930	8949	SEQ ID NO: 4236	attccagcttccaccatc	8338	8357	5
SEQ ID NO: 2908	ctttggactgtccaataag	8986	9005	SEQ ID NO: 4237	cttatgggatttctcaatg	11167	11186	5
SEQ ID NO: 2909	gcatctcacaacaatgaag	9260	9279	SEQ ID NO: 4238	cttcatctgtcatgtatgc	10227	10246	5
SEQ ID NO: 2910	cacaacaatgtgaagggaat	9265	9284	SEQ ID NO: 4239	atttccctgtgaattgtatg	11488	11507	5
SEQ ID NO: 2911	tccaaatttctctctgtgga	9415	9434	SEQ ID NO: 4240	tccatcacaaattctttgg	9671	9690	5
SEQ ID NO: 2912	caaaatttctctctgtgga	9416	9435	SEQ ID NO: 4241	tccatcacaaattctttgg	9670	9689	5
SEQ ID NO: 2913	ctctgtggaacaacagaga	9425	9444	SEQ ID NO: 4242	tctcaagagttacagcaga	13229	13248	5
SEQ ID NO: 2914	ctctgtggaacaacagaga	9426	9445	SEQ ID NO: 4243	tctcaagagttacagcag	13228	13247	5
SEQ ID NO: 2915	agaaacattatggaggcca	9441	9460	SEQ ID NO: 4244	ggggctggccccagattct	8909	8928	5
SEQ ID NO: 2916	agaaacaaattctgtatttc	9476	9495	SEQ ID NO: 4245	gaaattctcaatttattct	13821	13840	5
SEQ ID NO: 2917	ttctctcatgtggaaaaa	9565	9584	SEQ ID NO: 4246	tttttgcaagtaagaaaa	14021	14040	5
SEQ ID NO: 2918	tcagagcatcaaatcttt	9712	9731	SEQ ID NO: 4247	aaagaacatcagagcttga	14033	14052	5
SEQ ID NO: 2919	cagaacaatgcattagat	9751	9770	SEQ ID NO: 4248	atctatgacctctctctg	5633	5652	5
SEQ ID NO: 2920	tacacattatctgacct	10001	10020	SEQ ID NO: 4249	atgtgagttcttattgtga	14089	14108	5
SEQ ID NO: 2921	attcagatattgtctctca	10194	10213	SEQ ID NO: 4250	tgagaactcagagctgaact	4807	4826	5
SEQ ID NO: 2922	ggagggtagtcataacagt	10336	10355	SEQ ID NO: 4251	actgtgtgcacaaacctcc	2734	2753	5
SEQ ID NO: 2923	caaaagccgaattccaat	10404	10423	SEQ ID NO: 4252	atgtgaagtaacctattttg	8366	8385	5
SEQ ID NO: 2924	aaagccgaattccaatt	10405	10424	SEQ ID NO: 4253	gaattgaagtaacctttt	8365	8384	5
SEQ ID NO: 2925	tcaagcaagacctaattg	10436	10455	SEQ ID NO: 4254	attatggccctgtgtgaa	13256	13275	5
SEQ ID NO: 2926	ctcttacttttcaatga	10578	10597	SEQ ID NO: 4255	tcaaaagaagcccaagg	12947	12966	5
SEQ ID NO: 2927	tgaggccaacttactgt	10663	10682	SEQ ID NO: 4256	caagcatctgattgactca	12676	12695	5
SEQ ID NO: 2928	cacttactgtgaattccaag	10672	10691	SEQ ID NO: 4257	cttgaaacaaagtcagtg	6008	6027	5
SEQ ID NO: 2929	gaagtaagaagaattttg	10751	10770	SEQ ID NO: 4258	caaaaacattttcaactc	5287	5306	5
SEQ ID NO: 2930	ctggaaactctctcatgtg	10882	10901	SEQ ID NO: 4259	caatttaccagatcttcagg	11372	11391	5
SEQ ID NO: 2931	agctgtgatgaaccacag	11184	11203	SEQ ID NO: 4260	cttgattccacatgcagct	11855	11874	5
SEQ ID NO: 2932	aaatttccgtgaagtgat	11485	11504	SEQ ID NO: 4261	tatcatctccgtgttaatt	6765	6784	5
SEQ ID NO: 2933	cagatggcattgtctgttt	11613	11632	SEQ ID NO: 4262	aaagctgagaagaatactg	12424	12443	5
SEQ ID NO: 2934	agatgtgcattgtctgtttg	11614	11633	SEQ ID NO: 4263	caaaagctgagaagaatact	12423	12442	5
SEQ ID NO: 2935	gttgtaaaagctctgtgat	11842	11861	SEQ ID NO: 4264	atccaaagatgatacaaca	13103	13122	5
SEQ ID NO: 2936	catcttcaaaactcagtg	12229	12248	SEQ ID NO: 4265	caactctctgttactatgt	13631	13650	5
SEQ ID NO: 2937	aaagatttatacaaaagag	12938	12957	SEQ ID NO: 4266	cttcaatttattctcttt	13826	13845	5
SEQ ID NO: 2938	atttttccaactaatgaag	13034	13053	SEQ ID NO: 4267	cttcaagacttaaaaaat	8014	8033	5
SEQ ID NO: 2939	aattatatacagaatgaga	13097	13116	SEQ ID NO: 4268	tctcttcttccatgaatt	10479	10498	5
SEQ ID NO: 2940	tcaagaaagcttctcaaga	13218	13237	SEQ ID NO: 4269	cttctaagttacattgaa	13183	13202	5
SEQ ID NO: 2941	ttagcaatttctgacag	13437	13456	SEQ ID NO: 4270	ctgtgaagaatttatcaa	12932	12951	5
SEQ ID NO: 2942	ctgatatacatcacaggat	13712	13731	SEQ ID NO: 4271	actcaatggtaaatcag	7465	7484	5
SEQ ID NO: 2943	acatcacagtgattctgaa	13719	13738	SEQ ID NO: 4272	tttcaagaagtaagcaattg	7239	7258	5
SEQ ID NO: 2944	actgcctatattgataaaa	13882	13901	SEQ ID NO: 4273	ttttggcaagctatacagt	8380	8399	5
SEQ ID NO: 2945	agagatggcattttttgcaa	14011	14030	SEQ ID NO: 4274	ttgcaagcaagcttttct	3013	3032	5
SEQ ID NO: 2946	ttttttgcaagtaagaa	14020	14039	SEQ ID NO: 4275	tctctctatgggaaaaa	9566	9585	5

SEQ ID NO:	2947	tcagaaactcaagcttca	1627	1646	SEQ ID NO:	4276	gaatgctgtgttttga	8641	88603	4
SEQ ID NO:	2948	agttagtgaaagaagttct	1958	1975	SEQ ID NO:	4277	agaatctgtaccaggaact	12564	125833	4
SEQ ID NO:	2949	atttaccagctcgacaagt	5435	5454	SEQ ID NO:	4278	acttcagagaaatacaaat	11409	114283	4
SEQ ID NO:	2950	gattatctgaattcatcca	6488	6507	SEQ ID NO:	4279	gaacaacatgaacaaatc	7429	74483	4
SEQ ID NO:	2951	gtgacctctcgtgtgtcg	26	45	SEQ ID NO:	4280	cagctggagcagcagagac	6039	60582	4
SEQ ID NO:	2952	attcaagcaccctcgggaag	253	272	SEQ ID NO:	4281	ctctataagttcaattgaat	13184	132032	4
SEQ ID NO:	2953	gacgtcgatgaagaagatt	316	335	SEQ ID NO:	4282	acttcccaactctcaatgc	13415	134342	4
SEQ ID NO:	2954	gtctgacagcctccag	483	502	SEQ ID NO:	4283	ctgtggcagctgtatagcaa	5889	59082	4
SEQ ID NO:	2955	agaagatgaacctacta	555	574	SEQ ID NO:	4284	aaagtagttcaattct	10498	105172	4
SEQ ID NO:	2956	gaagactctcaggaact	1095	1114	SEQ ID NO:	4285	agttcaatgaattattca	13191	132102	4
SEQ ID NO:	2957	ttctctctgtccacagctg	1210	1229	SEQ ID NO:	4286	cagccagcagcattgagat	9237	92562	4
SEQ ID NO:	2958	ctctctgtccacagctga	1211	1230	SEQ ID NO:	4287	tcagccagcagcatttgaga	9236	92552	4
SEQ ID NO:	2959	tgagggtgcagcccccac	1231	1250	SEQ ID NO:	4288	gatgggaaagccgcoctca	5216	52352	4
SEQ ID NO:	2960	ccagaactcaagctcttcaa	1628	1647	SEQ ID NO:	4289	gttgaagcagaaacctctg	5915	59342	4
SEQ ID NO:	2961	ctgaaaaagttagtgaaag	1946	1968	SEQ ID NO:	4290	cttccggtgggaattatcag	10631	106502	4
SEQ ID NO:	2962	ttttccagacagtgta	2246	2265	SEQ ID NO:	4291	tgacaggcatttgaaaaa	9730	97492	4
SEQ ID NO:	2963	ttttccagacagtgta	2247	2266	SEQ ID NO:	4292	gtacaggcatttgaaaaa	9725	97482	4
SEQ ID NO:	2964	cattcagaacagaaatt	3403	3422	SEQ ID NO:	4293	aaattccaatttgaagaatg	10414	104332	4
SEQ ID NO:	2965	tgagagagaattgaatt	3628	3647	SEQ ID NO:	4294	aaatgtcagctctgttca	10902	109212	4
SEQ ID NO:	2966	ttgaatggaacacagga	3844	3863	SEQ ID NO:	4295	ggcagtttgaaaaaacaaa	11815	118342	4
SEQ ID NO:	2967	tctagattgtaaatca	4407	4426	SEQ ID NO:	4296	gtacatgttgataaagaa	7377	73962	4
SEQ ID NO:	2968	gattgttaatacaattca	4412	4431	SEQ ID NO:	4297	tgagtagaccacaacatc	7162	71812	4
SEQ ID NO:	2969	gcaagaccaccttgaa	5083	5102	SEQ ID NO:	4298	cttcaggttccatcgttga	11384	114032	4
SEQ ID NO:	2970	taagctctcaaalgat	5325	5344	SEQ ID NO:	4299	agtttgataaagaattaa	7382	74012	4
SEQ ID NO:	2971	caatttaacaacatgaat	6074	6093	SEQ ID NO:	4300	attcaactgcctatattg	13876	138952	4
SEQ ID NO:	2972	gaatacagccaggacttg	6088	6107	SEQ ID NO:	4301	caagagccacacggtctca	10687	107062	4
SEQ ID NO:	2973	ctcaaatattgataatt	6421	6440	SEQ ID NO:	4302	aaatttccctgaattgat	11488	115052	4
SEQ ID NO:	2974	tgagcatgtcaaacactt	7059	7078	SEQ ID NO:	4303	aaagtaagtcctaggttcaa	9381	94002	4
SEQ ID NO:	2975	tgagagagactattcagaa	7227	7246	SEQ ID NO:	4304	ttctctcagcagaaataatca	13446	134652	4
SEQ ID NO:	2976	ttcaggctcttcagaaagc	7829	7848	SEQ ID NO:	4305	gcttgcataactcctctga	12312	123312	4
SEQ ID NO:	2977	ttccacaaattgaacatcc	8787	8806	SEQ ID NO:	4306	gggaactaccagaagtgtga	12533	125522	4
SEQ ID NO:	2978	gaataccaatgtcgaact	10167	10186	SEQ ID NO:	4307	agtttcaatgaattattca	13191	132102	4
SEQ ID NO:	2979	tgaaactaagatgatc	12898	12917	SEQ ID NO:	4308	gattactatgaataattca	13640	136592	4
SEQ ID NO:	2980	tgacctgttccattcaaaa	13680	13699	SEQ ID NO:	4309	ttttaaagaattcttcaa	13813	138322	4
SEQ ID NO:	2981	gggtgagtgoccttctcg	19	38	SEQ ID NO:	4310	cgagccagccagccagccc	84	1031	4
SEQ ID NO:	2982	ggctgagtgoccttctcg	20	39	SEQ ID NO:	4311	ccagggccagccagccagcc	83	1021	4
SEQ ID NO:	2983	gtgagtgoccttctcggt	22	41	SEQ ID NO:	4312	aacgctgtcgaattcag	11557	115761	4
SEQ ID NO:	2984	ttctgtgtctgctgcgtga	33	52	SEQ ID NO:	4313	tcagctgacctcatcgaga	2168	21871	4
SEQ ID NO:	2985	gagccgagcagccagggagc	90	109	SEQ ID NO:	4314	gctctgcagcttctctg	376	3951	4
SEQ ID NO:	2986	ctgggctgctcgtcgtctg	151	170	SEQ ID NO:	4315	cagcagacacatttcaagc	4252	42711	4
SEQ ID NO:	2987	gtctgctgctggcgccag	177	196	SEQ ID NO:	4316	ctggaatgaaccaccagca	11186	112051	4
SEQ ID NO:	2988	ctgtgtctgtccaaaaggt	227	246	SEQ ID NO:	4317	catctgaagaccagccag	388	4071	4
SEQ ID NO:	2989	ctgagagttccagtgaggt	291	310	SEQ ID NO:	4318	actaccctgtgacacattg	3391	34101	4
SEQ ID NO:	2990	tccagtggtgagttccctggga	298	318	SEQ ID NO:	4319	ttcccgagccagaggtctga	2683	27021	4
SEQ ID NO:	2991	gggttgagctgaggttcc	354	373	SEQ ID NO:	4320	ggaaacctctccctcaact	4736	47551	4
SEQ ID NO:	2992	tgagctggaggttccag	358	377	SEQ ID NO:	4321	ctggaggagctgatgtctca	9171	91901	4

SEQ ID NO: 2993	ctgcagcttcactcgtgaa	378	397	SEQ ID NO: 4322	ttcaataataatcggcaga	3269	3288	4
SEQ ID NO: 2994	gccagtcaccctdgaaga	402	421	SEQ ID NO: 4323	cttcctcgtgtaagtc	5802	5821	4
SEQ ID NO: 2995	ctcagaggatttgtcgca	472	491	SEQ ID NO: 4324	ttcagaagaatttttagag	6348	6367	4
SEQ ID NO: 2996	aggatagagctcaagctgg	500	519	SEQ ID NO: 4325	ccagtttccggggaaacct	12724	12743	4
SEQ ID NO: 2997	tcctttaccctgggaaga	543	562	SEQ ID NO: 4326	ctttttgggaagaagga	2227	2246	4
SEQ ID NO: 2998	catcaagggggcatcatt	583	602	SEQ ID NO: 4327	taagtgtcaagttccgtgtg	2285	2304	4
SEQ ID NO: 2999	ctcgtgttcccccagagac	609	628	SEQ ID NO: 4328	gtctcgtcaactcagaagga	13996	14015	4
SEQ ID NO: 3000	gagcaagccaaagcatgtt	630	649	SEQ ID NO: 4329	taacaaataaaggagctct	14080	14099	4
SEQ ID NO: 3001	aagcaagtggtttctcgg	638	657	SEQ ID NO: 4330	ccagagcaggctcagctt	11050	11069	4
SEQ ID NO: 3002	ctcgatactcgttgatgga	652	671	SEQ ID NO: 4331	ttcaatgtccattatcaga	11364	11383	4
SEQ ID NO: 3003	ccactcatcttaccgtcaa	678	697	SEQ ID NO: 4332	ttgattttaacaaaagttg	6825	6844	4
SEQ ID NO: 3004	agggaaggcgaatgtgcaa	701	720	SEQ ID NO: 4333	ttgcaagcaagcttctct	3013	3032	4
SEQ ID NO: 3005	gcaattgtggcaacagaat	708	727	SEQ ID NO: 4334	taattccataccocgtttgc	3488	3507	4
SEQ ID NO: 3006	caatgtggcaacagaata	709	728	SEQ ID NO: 4335	ttatctctttttlcaattg	13834	13853	4
SEQ ID NO: 3007	ggcaacagaaatccac	714	733	SEQ ID NO: 4336	gtggcttccattatgtcca	1895	1914	4
SEQ ID NO: 3008	agagactctggggcagcttg	737	756	SEQ ID NO: 4337	caatttcaatttggctct	2938	2957	4
SEQ ID NO: 3009	gtgatgtcttcaacagca	752	771	SEQ ID NO: 4338	gtggaaagccgcctcaaca	5218	5237	4
SEQ ID NO: 3010	gtgatgtcttcaacagca	753	772	SEQ ID NO: 4339	atgggaaggaagccctcac	5217	5236	4
SEQ ID NO: 3011	gagccactgtctctctac	784	803	SEQ ID NO: 4340	gatgtcgaacagtagctg	8152	8171	4
SEQ ID NO: 3012	gctctcatcaaaaggcatga	794	813	SEQ ID NO: 4341	ttcataacagctactgtgagc	10345	10364	4
SEQ ID NO: 3013	cttgttcaactcgtgatcag	819	838	SEQ ID NO: 4342	gtcagtggttggtttacaag	12453	12472	4
SEQ ID NO: 3014	ctgttcaactcgtgatcag	820	839	SEQ ID NO: 4343	gtcagtggttggtttacaag	12452	12471	4
SEQ ID NO: 3015	agccatcttgaaggagca	892	911	SEQ ID NO: 4344	ttcaatgtgctcattggct	3813	3832	4
SEQ ID NO: 3016	gcaatctgcaaggagcaac	893	912	SEQ ID NO: 4345	gttgcaatgagctcatgcg	3812	3831	4
SEQ ID NO: 3017	cttctcgtcttctctac	916	935	SEQ ID NO: 4346	gttaggaataaatggagaag	9461	9480	4
SEQ ID NO: 3018	ctttctctacaagaataa	924	943	SEQ ID NO: 4347	ttatgtcgtgaatccaaag	13656	13675	4
SEQ ID NO: 3019	gatacaacgcgcgtcttt	997	1016	SEQ ID NO: 4348	aaagccatcatcgtgatc	1689	1688	4
SEQ ID NO: 3020	atcaacagcgcgtctttt	998	1017	SEQ ID NO: 4349	caagccatcatcgtgatc	1688	1687	4
SEQ ID NO: 3021	acagcgcgtcttttggga	1002	1021	SEQ ID NO: 4350	ttcaaatctctttgtcgt	9675	9694	4
SEQ ID NO: 3022	agatgggcgtctgcattg	1031	1050	SEQ ID NO: 4351	caaatagaagggaatctt	2077	2096	4
SEQ ID NO: 3023	ttttgaagactctccag	1090	1109	SEQ ID NO: 4352	ctggtaactctttaaaca	5495	5514	4
SEQ ID NO: 3024	ttgaagactctccaggaa	1094	1113	SEQ ID NO: 4353	gttcaatgaatttatcaac	13192	13211	4
SEQ ID NO: 3025	aactgaaataaacatcat	1110	1129	SEQ ID NO: 4354	atgtgcatctttttgcaagt	14014	14033	4
SEQ ID NO: 3026	tgaaaaaaataaccatct	1112	1131	SEQ ID NO: 4355	agatgtatgggcagcttag	4572	4591	4
SEQ ID NO: 3027	aaaaataaccatctctgag	1117	1136	SEQ ID NO: 4356	ctcaagaatgactttttt	2578	2597	4
SEQ ID NO: 3028	gtgcaaaatatacagaga	1132	1151	SEQ ID NO: 4357	ttccagagataaaaaccta	12209	12228	4
SEQ ID NO: 3029	caataagctgtgtactgag	1162	1181	SEQ ID NO: 4358	ctcagatcaaaagttaattg	12273	12292	4
SEQ ID NO: 3030	tactgactgagaggcctc	1174	1193	SEQ ID NO: 4359	gagggtagtcaataacagta	10337	10356	4
SEQ ID NO: 3031	gctcagtgatgaagcagt	1188	1207	SEQ ID NO: 4360	actgttgactcagggaagc	12580	12599	4
SEQ ID NO: 3032	agtcacatctctcttgcca	1204	1223	SEQ ID NO: 4361	gtggccatagatctgagct	8866	8885	4
SEQ ID NO: 3033	atctctctgtggcaacgtg	1210	1229	SEQ ID NO: 4362	gagctgacctcatcgagat	2169	2188	4
SEQ ID NO: 3034	ctctctctgtggcaacgtg	1211	1230	SEQ ID NO: 4363	tcagctgacctcatcgaga	2168	2187	4
SEQ ID NO: 3035	gtccacagctgatgaggt	1218	1237	SEQ ID NO: 4364	actgtgcaccaagctgtgc	13963	13982	4
SEQ ID NO: 3036	gccacagctgatgaggtg	1219	1238	SEQ ID NO: 4365	accacaaataacccaatgc	11248	11267	4
SEQ ID NO: 3037	tcacttacaagccttgtt	1248	1267	SEQ ID NO: 4366	accagatgtcgtcaacagta	8148	8167	4
SEQ ID NO: 3038	ccctctgatagatgtgt	1332	1351	SEQ ID NO: 4367	accacttacagctagagac	10824	10843	4
SEQ ID NO: 3039	gtcactcatctgggtggcc	1349	1368	SEQ ID NO: 4368	gggcagacttaagttgtgac	3439	3458	4

SEQ ID NO: 3040	ctgtgtatgcgtcagacca	1440	1459	SEQ ID NO: 3369	ggcttgcgtacacaaagg	5586	5605	1	4
SEQ ID NO: 3041	gacaaaccctacaggagacc	1480	1499	SEQ ID NO: 3370	ggcttcttatgatattgc	12355	12374	1	4
SEQ ID NO: 3042	gctaattaccgtatgtgaa	1516	1535	SEQ ID NO: 3371	ttcccaaaagcagtcagca	9938	9957	1	4
SEQ ID NO: 3043	gactcgcactgggagtgaa	1546	1565	SEQ ID NO: 3372	ttcaggttcgatctcaagtc	10917	10936	1	4
SEQ ID NO: 3044	actcgcactggggatgaaga	1548	1567	SEQ ID NO: 3373	ctgtgttaacacaaagctcgt	6007	6026	1	4
SEQ ID NO: 3045	atgaagattacacctattt	1560	1579	SEQ ID NO: 3374	aaatgaagtaaaagatcat	8118	8137	1	4
SEQ ID NO: 3046	accatggagcagttacaactc	1610	1629	SEQ ID NO: 3375	ggatgaataccaaacttggt	9024	9043	1	4
SEQ ID NO: 3047	gcagtttaactcagaactc	1618	1637	SEQ ID NO: 3376	ggatgtactgcaaaagctcgt	13727	13746	1	4
SEQ ID NO: 3048	cagaactcaagtcttcaat	1629	1648	SEQ ID NO: 3377	attggatatacaagatctg	1933	1952	1	4
SEQ ID NO: 3049	cagctctcgggaaaattg	1703	1722	SEQ ID NO: 3378	ccatgcacctcagctcctcg	2485	2504	1	4
SEQ ID NO: 3050	ccaggaggttctctcttc	1738	1757	SEQ ID NO: 3379	gtcgaatacaatgctctgg	5519	5538	1	4
SEQ ID NO: 3051	ggttcttctcagacittc	1744	1763	SEQ ID NO: 3380	gaaaaactgtgaaacacc	4439	4458	1	4
SEQ ID NO: 3052	ttctctgatgatgtcttct	1759	1778	SEQ ID NO: 3381	agaatccagatacaagaaa	6893	6912	1	4
SEQ ID NO: 3053	ggagataagcagctggctg	1781	1800	SEQ ID NO: 3382	cagactgctgactgttctcc	9952	9971	1	4
SEQ ID NO: 3054	gtcgtctatcttatgttga	1796	1815	SEQ ID NO: 3383	ccaatatacaaaagcccagc	12045	12064	1	4
SEQ ID NO: 3055	actttgggcttcccatat	1890	1909	SEQ ID NO: 3384	gatctctggaactctgaagt	10737	10756	1	4
SEQ ID NO: 3056	gccaatatcttgaactcag	1910	1929	SEQ ID NO: 3385	ctgaactcagaaggatggc	14000	14019	1	4
SEQ ID NO: 3057	aatatcttgaactcagaag	1913	1932	SEQ ID NO: 3386	cttccaattctgaaatatt	13378	13397	1	4
SEQ ID NO: 3058	ctcagaagaattgtgatalc	1924	1943	SEQ ID NO: 3387	gtataaagaattactttgaag	7273	7292	1	4
SEQ ID NO: 3059	agaattgtatataccaaga	1929	1948	SEQ ID NO: 3388	cttcaattattcttctt	13825	13844	1	4
SEQ ID NO: 3060	ggaattggatataccaagat	1930	1949	SEQ ID NO: 3389	attcttccaattattctctt	13824	13843	1	4
SEQ ID NO: 3061	ggagatatccaagcttgaa	1935	1954	SEQ ID NO: 3390	ttcatacaccagaattcca	8325	8344	1	4
SEQ ID NO: 3062	gtatccaagatctgaaaaa	1938	1957	SEQ ID NO: 3391	tttttaaccagctcagat	10185	10204	1	4
SEQ ID NO: 3063	atccaagatctgaaaaag	1939	1958	SEQ ID NO: 3392	cttttaaccagctcagata	10184	10203	1	4
SEQ ID NO: 3064	caagatctgaaaaagttgag	1943	1962	SEQ ID NO: 3393	ctaaattcccaattgtcttg	4973	4992	1	4
SEQ ID NO: 3065	aagatctgaaaaagttagt	1944	1963	SEQ ID NO: 3394	actaaattcccaattgtctt	4972	4991	1	4
SEQ ID NO: 3066	gaaaaagttagtgaaaga	1950	1969	SEQ ID NO: 3395	ctttctcgtgggaattatca	10630	10649	1	4
SEQ ID NO: 3067	ttccaactgtcatggacttc	1990	2009	SEQ ID NO: 3396	gaagcacatatagaactgga	13945	13964	1	4
SEQ ID NO: 3068	tcagaaaaattctctggaa	2007	2026	SEQ ID NO: 3397	ttcttttaacaattcctga	9501	9520	1	4
SEQ ID NO: 3069	ttccatcaacttgaccagc	2052	2071	SEQ ID NO: 3398	gctgcacatagggaatggaa	8441	8460	1	4
SEQ ID NO: 3070	ccagcctcagccaaaata	2065	2084	SEQ ID NO: 3399	attctatccaagaattggg	7820	7839	1	4
SEQ ID NO: 3071	ggcctcagccaaaatagaa	2068	2087	SEQ ID NO: 3400	ttctatccaagaattgggct	7822	7841	1	4
SEQ ID NO: 3072	gtctctatatattgatccaa	2091	2110	SEQ ID NO: 3401	ttgcaaaaaaagacagat	11821	11840	1	4
SEQ ID NO: 3073	ttctatatattgatccaat	2092	2111	SEQ ID NO: 3402	attttttgcagtttaaga	14019	14038	1	4
SEQ ID NO: 3074	cttctaaagaagaacatgc	2117	2136	SEQ ID NO: 3403	gcatggcattatgatgaag	3614	3633	1	4
SEQ ID NO: 3075	ctaaagaagcatctcgtaa	2121	2140	SEQ ID NO: 3404	ttcaggggtgtggagttgag	5694	5713	1	4
SEQ ID NO: 3076	baaagaagaacatgctgaa	2122	2141	SEQ ID NO: 3405	ttcttaaacattctcttata	9490	9509	1	4
SEQ ID NO: 3077	gagattggcttggaaaggaa	2183	2202	SEQ ID NO: 3406	ttccctccaattagttctc	11709	11728	1	4
SEQ ID NO: 3078	ctttgagccaaactgtgaa	2206	2225	SEQ ID NO: 3407	ttccatgaccaagaagaaa	11068	11087	1	4
SEQ ID NO: 3079	cagcagtggtcaacaagc	2253	2272	SEQ ID NO: 3408	gtctactctggaagaaactcg	6142	6161	1	4
SEQ ID NO: 3080	cagtggtcaacaagaattgt	2257	2276	SEQ ID NO: 3409	caaatctctgatacactcg	9857	9876	1	4
SEQ ID NO: 3081	gtgtgtcaacaagaattgt	2258	2277	SEQ ID NO: 3410	acaagaataatgctacact	4359	4378	1	4
SEQ ID NO: 3082	ctgatgtgtctcttaagtt	2298	2317	SEQ ID NO: 3411	taactcgggaaactctctag	3333	3352	1	4
SEQ ID NO: 3083	gtatgtgtctcttaagtc	2299	2318	SEQ ID NO: 3412	gaactcgggcaacagatca	8831	8850	1	4
SEQ ID NO: 3084	aaacatgagcaggatattg	2351	2370	SEQ ID NO: 3413	ccatgatctacatttttttt	6796	6815	1	4
SEQ ID NO: 3085	gaagctgattaaagattgt	2395	2414	SEQ ID NO: 3414	caaaaacattttcaacttc	5287	5306	1	4
SEQ ID NO: 3086	aaagattgaaatccaag	2405	2424	SEQ ID NO: 3415	cttaagtctcagcatctt	7614	7633	1	4

SEQ ID NO: 3087	patgggtccgcactc	2518	2537	SEQ ID NO: 4416	cagattgaggtatcc	7983	8002	4
SEQ ID NO: 3088	ggatcccccagatg	2540	2559	SEQ ID NO: 4417	caatcacagctgattcc	9083	9102	4
SEQ ID NO: 3089	ttttctactacatctc	2593	2612	SEQ ID NO: 4418	gaagtgatcagtgcaaaa	10382	10401	4
SEQ ID NO: 3090	cttcaactacatctcatg	2596	2615	SEQ ID NO: 4419	catggtcattatgaaag	3615	3634	4
SEQ ID NO: 3091	acatctctcatggaatg	2603	2622	SEQ ID NO: 4420	catatgagggccatgla	9445	9464	4
SEQ ID NO: 3092	ttcatgagaaatgctt	2604	2628	SEQ ID NO: 4421	caaaatcaactttaatga	6607	6626	4
SEQ ID NO: 3093	catggagaatgcttga	2610	2629	SEQ ID NO: 4422	caacacaaatcttaagla	13116	13135	4
SEQ ID NO: 3094	ttgaactccacatggag	2624	2643	SEQ ID NO: 4423	ctccacaggaactttcaaa	9842	9861	4
SEQ ID NO: 3095	ttgaactccacatggagc	2625	2644	SEQ ID NO: 4424	gtcccccaggaactttcaa	9841	9860	4
SEQ ID NO: 3096	gaactccacatggagct	2626	2645	SEQ ID NO: 4425	agctccacaggaactttca	9840	9859	4
SEQ ID NO: 3097	catggagctggtatcag	2635	2654	SEQ ID NO: 4426	ctgtttctgagctccag	9344	9363	4
SEQ ID NO: 3098	agctgagctggtatcag	2636	2655	SEQ ID NO: 4427	actgtttctgagctccag	9343	9362	4
SEQ ID NO: 3099	agttgcaaatatctcatc	2652	2671	SEQ ID NO: 4428	gatgatgccaaaatcaact	6599	6618	4
SEQ ID NO: 3100	gttgcaaatatctcatct	2653	2672	SEQ ID NO: 4429	gatgatgccaaaatcaac	6598	6617	4
SEQ ID NO: 3101	aaatattctcatctggat	2658	2677	SEQ ID NO: 4430	actcagaagatgcatatt	14004	14023	4
SEQ ID NO: 3102	aaacatcggaagtagccaa	2703	2722	SEQ ID NO: 4431	ttgtacagagagcttca	7600	7619	4
SEQ ID NO: 3103	ggctgaactctgggcaaaa	2728	2747	SEQ ID NO: 4432	ttttctttcagccagoc	9228	9247	4
SEQ ID NO: 3104	cttgagatttgacaaat	2758	2777	SEQ ID NO: 4433	attttcaagcaaatgcaca	8538	8557	4
SEQ ID NO: 3105	ttgttgacaaatagggcat	2766	2785	SEQ ID NO: 4434	atgctctacacttaccaca	9521	9540	4
SEQ ID NO: 3106	atgaacacaaactttctc	2819	2838	SEQ ID NO: 4435	ggagctggaatttatcat	2877	2896	4
SEQ ID NO: 3107	cttcacagatgctgggtct	2833	2852	SEQ ID NO: 4436	cagagctatcacgtgggaag	5235	5254	4
SEQ ID NO: 3108	gatctgggtctggagctc	2840	2859	SEQ ID NO: 4437	gagcttactggaagaactc	6140	6159	4
SEQ ID NO: 3109	ctataaagcttggaagctg	2866	2885	SEQ ID NO: 4438	cagccctccacgcgctag	12120	12139	4
SEQ ID NO: 3110	agctgggaagctggaatt	2872	2891	SEQ ID NO: 4439	aaactgttaattacagct	5463	5482	4
SEQ ID NO: 3111	ccagatttagagctggaact	3114	3133	SEQ ID NO: 4440	agtttccggggaactcgt	12726	12745	4
SEQ ID NO: 3112	tgataccctgaagtttga	3208	3227	SEQ ID NO: 4441	ttacagattctgaaaatct	8393	8412	4
SEQ ID NO: 3113	ctgaggtaccatcgacatt	3252	3271	SEQ ID NO: 4442	aatgagctctatgcttca	3817	3836	4
SEQ ID NO: 3114	gtccagtggaagtcacaa	3297	3316	SEQ ID NO: 4443	attttgagaggaatgcaca	6357	6376	4
SEQ ID NO: 3115	aaatccggaatttgatgt	3313	3332	SEQ ID NO: 4444	aaacatgaatcaacaatt	8938	8957	4
SEQ ID NO: 3116	ttccggaatttgatgta	3315	3334	SEQ ID NO: 4445	caaaacagagctcaggaa	13207	13226	4
SEQ ID NO: 3117	ggaaacaaatctccagatt	3337	3356	SEQ ID NO: 4446	aaactgtacaaatggtccg	4211	4230	4
SEQ ID NO: 3118	ctccagagttaatgatga	3345	3364	SEQ ID NO: 4447	ttcatatgttgatcagga	7593	7612	4
SEQ ID NO: 3119	ttcaaccctcagacttcaga	3392	3411	SEQ ID NO: 4448	ctgcagacaaatgctag	12439	12458	4
SEQ ID NO: 3120	ctttcagaaacagaataatt	3403	3422	SEQ ID NO: 4449	aatgtactttatgaaatg	8104	8123	4
SEQ ID NO: 3121	actgagtgctgacctatg	3422	3441	SEQ ID NO: 4450	ccatgcaagtcagccaggt	10924	10943	4
SEQ ID NO: 3122	tatttccataaccogt	3486	3505	SEQ ID NO: 4451	aaactgcotatgtatgata	13880	13899	4
SEQ ID NO: 3123	gtttgcaagcagaagccag	3501	3520	SEQ ID NO: 4452	ctggaactctttcaaaac	5408	5427	4
SEQ ID NO: 3124	ttgcaagcagaagccagca	3502	3521	SEQ ID NO: 4453	ctgggtgtgctacagcaaa	5272	5291	4
SEQ ID NO: 3125	ttgcaagcagaagccagaa	3503	3522	SEQ ID NO: 4454	ttctgggtgtgctacagcaa	5271	5290	4
SEQ ID NO: 3126	ctgtcttctccaaatggact	3554	3573	SEQ ID NO: 4455	agtcagaagattgctggcag	4567	4586	4
SEQ ID NO: 3127	gtctacagcttatgtctcc	3577	3596	SEQ ID NO: 4456	ggaggctttaagttcagca	7609	7628	4
SEQ ID NO: 3128	acagcttatgtctccacag	3581	3600	SEQ ID NO: 4457	ctgtatagcaaatctctgt	5897	5916	4
SEQ ID NO: 3129	ttccaagagggtgcatct	3600	3619	SEQ ID NO: 4458	catggaactttctggaaa	8877	8896	4
SEQ ID NO: 3130	ccaagaagggtgcatgcca	3603	3622	SEQ ID NO: 4459	gtccagcaagcaagttgt	9361	9380	4
SEQ ID NO: 3131	gtgcatgctattatgat	3611	3630	SEQ ID NO: 4460	catcttaacacactccac	8071	8090	4
SEQ ID NO: 3132	gtatgagaagaagattgaa	3625	3644	SEQ ID NO: 4461	ttcaacttctctgaaatca	7871	7890	4
SEQ ID NO: 3133	gaagagaagaattgattgt	3629	3648	SEQ ID NO: 4462	caaaacaaactttcaacttc	5287	5306	4

SEQ ID NO. 3134	gagaagatgaattgaat	3632	3651	SEQ ID NO. 4463	atcataatccaactctc	8278	8297	4
SEQ ID NO. 3135	ttgaalggaacacagcca	3644	3663	SEQ ID NO. 4464	ggcctttgtacacaaa	11236	11255	4
SEQ ID NO. 3136	gggacccaatgtatagacc	3658	3677	SEQ ID NO. 4465	gglaacccaagaagagcct	5691	5610	4
SEQ ID NO. 3137	caaaaaaatgactccaat	3676	3695	SEQ ID NO. 4466	atgaagtactacttttg	8366	8385	4
SEQ ID NO. 3138	aaaaaaatgctccaatt	3677	3696	SEQ ID NO. 4467	aatgaagtactactattt	8365	8384	4
SEQ ID NO. 3139	aaaaaatgactccaatt	3678	3697	SEQ ID NO. 4468	aaatcaatctctctttt	8406	8425	4
SEQ ID NO. 3140	cagagtccccaacagac	3760	3779	SEQ ID NO. 4469	gtctggggatccatctg	4090	4109	4
SEQ ID NO. 3141	aaataatagtccaalga	3803	3822	SEQ ID NO. 4470	tcaatgaatccaalgaatt	13186	13205	4
SEQ ID NO. 3142	tcaactccagaacatgg	3899	3918	SEQ ID NO. 4471	ccatlgaccagatgctgaa	8142	8161	4
SEQ ID NO. 3143	tgggagtgcagacactga	3915	3934	SEQ ID NO. 4472	tggaaatggcctggccca	8903	8922	4
SEQ ID NO. 3144	cagtttgaatagtgaatt	3994	4013	SEQ ID NO. 4473	aatcacactctctcactg	9541	9560	4
SEQ ID NO. 3145	gaaaatgagatctcttg	4000	4019	SEQ ID NO. 4474	caaaactaccacacatttc	13694	13713	4
SEQ ID NO. 3146	ttgccttttggggcaaa	4015	4034	SEQ ID NO. 4475	ttgagagggaatgcacaaa	6359	6378	4
SEQ ID NO. 3147	ctccagagatctaaagatg	4036	4055	SEQ ID NO. 4476	ctcataatgggtacaggag	7594	7613	4
SEQ ID NO. 3148	ctaaagaatgtatagact	4045	4064	SEQ ID NO. 4477	agctctctatgtctcctga	10033	10052	4
SEQ ID NO. 3149	ctgtggatctccatctgc	4092	4111	SEQ ID NO. 4478	ggcatttgaaaaaaacag	9735	9754	4
SEQ ID NO. 3150	atctgcctcatcagagtt	4104	4123	SEQ ID NO. 4479	aactctcaaacctcaagat	8556	8575	4
SEQ ID NO. 3151	ctcagagatgtccaagtc	4112	4131	SEQ ID NO. 4480	ggacattctctctagcaga	8215	8234	4
SEQ ID NO. 3152	gtctccactctttacatt	4128	4145	SEQ ID NO. 4481	aatgaatcagccaggact	6086	6105	4
SEQ ID NO. 3153	atttaccattccaagt	4133	4152	SEQ ID NO. 4482	actttgagaagaagaagt	8109	8128	4
SEQ ID NO. 3154	catctccaagtgtatcaa	4141	4160	SEQ ID NO. 4483	ttgaaggactcaggaagt	12009	12028	4
SEQ ID NO. 3155	accacatgaagctgactc	4284	4303	SEQ ID NO. 4484	gagtaaacccaagaacttgt	9024	9043	4
SEQ ID NO. 3156	ttctcaaatgtgcgaagg	4317	4336	SEQ ID NO. 4485	cccttaacaattctcgaaa	9503	9522	4
SEQ ID NO. 3157	ctgggaacaacatatga	4338	4357	SEQ ID NO. 4486	catctcgggtctttccag	11035	11054	4
SEQ ID NO. 3158	atcatgtgtgggtctctca	4378	4397	SEQ ID NO. 4487	tgaattacagaaaatgat	6565	6584	4
SEQ ID NO. 3159	catgtgatgggtctctaac	4380	4399	SEQ ID NO. 4488	cgttaggcacacgtggcctg	12133	12152	4
SEQ ID NO. 3160	tctagattcgaaatcaaa	4407	4426	SEQ ID NO. 4489	ttgatgatgtcttcaagaa	7308	7327	4
SEQ ID NO. 3161	tgggaccacagatgctg	4499	4518	SEQ ID NO. 4490	cagaattccagctcccca	8334	8353	4
SEQ ID NO. 3162	ctaacactggccggtctaac	4644	4663	SEQ ID NO. 4491	ttgaggtctatgatgtatg	6984	7003	4
SEQ ID NO. 3163	taacactggccggtcatt	4645	4664	SEQ ID NO. 4492	atggagctatgatgtata	6983	7002	4
SEQ ID NO. 3164	taacactggccggtctaatg	4646	4665	SEQ ID NO. 4493	catggagctatgatgtt	6982	7001	4
SEQ ID NO. 3165	ctggccggtcattgagaa	4650	4669	SEQ ID NO. 4494	ctctcatctgcgctacactg	12073	12092	4
SEQ ID NO. 3166	agataacagggaatata	4713	4732	SEQ ID NO. 4495	ctactctctctctccat	10210	10229	4
SEQ ID NO. 3167	tccttaactccacctctg	4745	4764	SEQ ID NO. 4496	cagatatatactcaggga	8184	8203	4
SEQ ID NO. 3168	agctgactttaaactctga	4818	4837	SEQ ID NO. 4497	tcaggctctcagaagctg	7930	7949	4
SEQ ID NO. 3169	ctgactttaaactctgaca	4820	4839	SEQ ID NO. 4498	gtcaagataaacactatcg	8740	8759	4
SEQ ID NO. 3170	caagatggatagaccttc	4873	4892	SEQ ID NO. 4499	gaagtagtacttgcacttc	6843	6862	4
SEQ ID NO. 3171	actggtcttgatgaatcatg	4909	4928	SEQ ID NO. 4500	ctgagctccagtgccaccg	9350	9369	4
SEQ ID NO. 3172	ctgtctgaatatcagctg	4913	4932	SEQ ID NO. 4501	tcagcaagtaacctgagaacg	8611	8630	4
SEQ ID NO. 3173	aatctccatgtctgtgaat	4976	4995	SEQ ID NO. 4502	actcagatacaaatgtaatt	12272	12291	4
SEQ ID NO. 3174	ggctctgagtaaatgct	4984	5003	SEQ ID NO. 4503	agcacagatcgaaaaacca	10809	10828	4
SEQ ID NO. 3175	cttgagttaaagtctgaca	4988	5007	SEQ ID NO. 4504	gtccctagaaaactcaag	10042	10061	4
SEQ ID NO. 3176	cttgagttaaagtctgacat	4989	5008	SEQ ID NO. 4505	atgtcctcagataaatctcaa	10041	10060	4
SEQ ID NO. 3177	gagttaaagtctgacatc	4990	5009	SEQ ID NO. 4506	gatggaaacctctcctca	4733	4752	4
SEQ ID NO. 3178	actgaagtgtagtctcct	5094	5113	SEQ ID NO. 4507	aggaactcagatcaaatg	12267	12286	4
SEQ ID NO. 3179	agtgtagtctcctggctg	5100	5119	SEQ ID NO. 4508	agcagccagtgccacact	12514	12533	4
SEQ ID NO. 3180	gtctgagaaatgactga	5114	5133	SEQ ID NO. 4509	tcagccaggttatagcac	7734	7753	4

SEQ ID NO: 3181	ctggggcatcatgaatt	5151	5170	SEQ ID NO: 4510	aattctgatiacaaccag	13579	13598	4
SEQ ID NO: 3182	atggcgccttcagggaaca	5178	5197	SEQ ID NO: 4511	tgcttttggaaagccat	8649	8668	4
SEQ ID NO: 3183	tcagtcctggagtggaaga	5207	5226	SEQ ID NO: 4512	cttgacagggcatlftgaa	9727	9746	4
SEQ ID NO: 3184	ccatgattctggggtgtga	5265	5284	SEQ ID NO: 4513	ctgagcacatcaaaagt	5838	5857	4
SEQ ID NO: 3185	aaacacatttcaactcoa	5289	5308	SEQ ID NO: 4514	ctgatgtttagagtgcttt	6993	7012	4
SEQ ID NO: 3186	cttaagcttcaaatgaca	5324	5343	SEQ ID NO: 4515	gtctccacaacaagttaag	7255	7274	4
SEQ ID NO: 3187	ttaagcttcaaatgacat	5325	5344	SEQ ID NO: 4516	gatgtctccacaagaattaa	7254	7273	4
SEQ ID NO: 3188	catgagtggtctatgctct	5341	5360	SEQ ID NO: 4517	agcatctgtgagctacagt	7624	7643	4
SEQ ID NO: 3189	gggtctcatatgctgaatt	5346	5365	SEQ ID NO: 4518	attttatcaaaaagaagccca	12942	12961	4
SEQ ID NO: 3190	actggactctcttcaaaa	5407	5426	SEQ ID NO: 4519	tttggcagaagctacagtt	8380	8399	4
SEQ ID NO: 3191	actctcttcaaaactga	5412	5431	SEQ ID NO: 4520	caattggggagagacaagt	6504	6523	4
SEQ ID NO: 3192	ctgacaagtttataaaga	5445	5464	SEQ ID NO: 4521	tgcttggtagtttatcag	9693	9712	4
SEQ ID NO: 3193	aagtttataagcaactgt	5450	5469	SEQ ID NO: 4522	cagctatgtagaaaaactt	4429	4448	4
SEQ ID NO: 3194	ctgttaatttacagctaca	5466	5485	SEQ ID NO: 4523	gtcatgggaaaagctacag	6388	6407	4
SEQ ID NO: 3195	ttaacagctacagccattt	5474	5493	SEQ ID NO: 4524	aatattgatcaattgttaa	6425	6444	4
SEQ ID NO: 3196	ctgtgttaactatttaaac	5494	5513	SEQ ID NO: 4525	gtttgaaaacaagaagcaga	11820	11839	4
SEQ ID NO: 3197	ttaaacagtgacctgaaa	5506	5525	SEQ ID NO: 4526	tttcatlftgaaagataaa	7032	7051	4
SEQ ID NO: 3198	ttaaacagtgacctgaatt	5507	5526	SEQ ID NO: 4527	atttcagcagaagacttaa	10434	10453	4
SEQ ID NO: 3199	cagtgacctgaataacaat	5512	5531	SEQ ID NO: 4528	attgtgtgtggagctactgt	6131	6150	4
SEQ ID NO: 3200	ctgggtgtgttaacttaaaa	5584	5603	SEQ ID NO: 4529	ttgtctgagaagaagccaca	10765	10784	4
SEQ ID NO: 3201	tatcagcagaactctaaag	5657	5676	SEQ ID NO: 4530	cttgcacatgtgtcataa	12764	12783	4
SEQ ID NO: 3202	gttgcagggtgtggagttt	5692	5711	SEQ ID NO: 4531	taaacacctaagaagtaaac	9014	9033	4
SEQ ID NO: 3203	attcagactcatctgcatlft	5775	5794	SEQ ID NO: 4532	aatgtctgacataggggaat	8437	8456	4
SEQ ID NO: 3204	ttcagactcaactgatttc	5776	5795	SEQ ID NO: 4533	gaatatattgaacttgaa	13312	13331	4
SEQ ID NO: 3205	tacaattggcaatgggaaa	5848	5867	SEQ ID NO: 4534	ttctcaaaagctggatgta	11176	11195	4
SEQ ID NO: 3206	gtctatagcaaatctctc	5896	5915	SEQ ID NO: 4535	caggttcacatgcaagtcagc	10919	10938	4
SEQ ID NO: 3207	tgacagacagggcaactgg	6043	6062	SEQ ID NO: 4536	ccagctctccacacatctca	8341	8360	4
SEQ ID NO: 3208	ggcactdggaaactcaaga	6053	6072	SEQ ID NO: 4537	ctctctgtgttcaactgoc	11221	11240	4
SEQ ID NO: 3209	gaatacagocaggactgt	6088	6107	SEQ ID NO: 4538	caagtaagtgtcagggttca	9380	9399	4
SEQ ID NO: 3210	gaatacagocaggactgtg	6089	6108	SEQ ID NO: 4539	ccacacatctactgaaatc	10668	10687	4
SEQ ID NO: 3211	ctggagcaactctgctgta	6147	6166	SEQ ID NO: 4540	cagaagaactacatctcag	7939	7958	4
SEQ ID NO: 3212	tttactcagtgagcccat	6201	6220	SEQ ID NO: 4541	atggactctctctggaaaa	8878	8897	4
SEQ ID NO: 3213	gatgagagatgcggttgag	6241	6260	SEQ ID NO: 4542	ctcatctctctcttctcatc	10209	10228	4
SEQ ID NO: 3214	aatgttgcttctgttaaa	6277	6296	SEQ ID NO: 4543	ctttcttaactlftgaatt	9064	9083	4
SEQ ID NO: 3215	ctttgttaagatgataaa	6285	6304	SEQ ID NO: 4544	ttatgaactlftgaagaatg	13318	13337	4
SEQ ID NO: 3216	tttgaagatgataaaaa	6287	6306	SEQ ID NO: 4545	tttccatattagatgcaaa	8421	8440	4
SEQ ID NO: 3217	ttccattaccctccatttt	6320	6339	SEQ ID NO: 4546	aaaattgatgatcttggga	10727	10746	4
SEQ ID NO: 3218	ctccattaccctccatttt	6321	6340	SEQ ID NO: 4547	aaaagggctcaggaatgg	8893	8912	4
SEQ ID NO: 3219	cttgcagaatattttgag	6346	6365	SEQ ID NO: 4548	ctcaattttgttttcaag	8528	8547	4
SEQ ID NO: 3220	agaatattttgagaggaat	6352	6371	SEQ ID NO: 4549	attccctccattgaattct	11708	11727	4
SEQ ID NO: 3221	attatgltgactgtgaaa	6380	6399	SEQ ID NO: 4550	tttcaagcagaagaactaat	10435	10454	4
SEQ ID NO: 3222	gaagcacatcaattatgat	6415	6434	SEQ ID NO: 4551	atcagttcagataaacttc	7999	8018	4
SEQ ID NO: 3223	acatcaatattgatcaatt	6420	6439	SEQ ID NO: 4552	aattccctgaagttgatgt	11487	11506	4
SEQ ID NO: 3224	gaaaaactccacagcagaac	6485	6484	SEQ ID NO: 4553	gtctctctctccacatttc	10060	10079	4
SEQ ID NO: 3225	ctgaattcttcaatttggg	6494	6513	SEQ ID NO: 4554	cccatctacagactlftca	11371	11390	4
SEQ ID NO: 3226	gaattcttcaatttggga	6495	6514	SEQ ID NO: 4555	cccatctacagactlftca	11370	11389	4
SEQ ID NO: 3227	taactgactctctcaaaa	6540	6559	SEQ ID NO: 4556	tttgaggatctcactgatt	7987	8006	4

SEQ ID NO: 3228	aaagatagataattacaga	6558	6577	SEQ ID NO: 4557	ctgctccctcaacttt	9050	9069	1	4
SEQ ID NO: 3229	ctcaacttaalgaaaac	6611	6630	SEQ ID NO: 4558	gtttatgaaatattgat	6811	6830	1	4
SEQ ID NO: 3230	gatttgaaaatgctatt	6694	6713	SEQ ID NO: 4559	aatatttgatgaataca	6716	6735	1	4
SEQ ID NO: 3231	atttgaaaatgctattgc	6696	6715	SEQ ID NO: 4560	gcagagacttaaggaaat	10441	10460	1	4
SEQ ID NO: 3232	attgcaatattattgatg	6710	6729	SEQ ID NO: 4561	catcacactgaataacat	10159	10178	1	4
SEQ ID NO: 3233	gaaaaattaaaagctctg	6737	6756	SEQ ID NO: 4562	caagagcttatggatttc	11161	11180	1	4
SEQ ID NO: 3234	actcatcatccgtgtaai	6762	6781	SEQ ID NO: 4563	tactacttgagaaatagt	7281	7300	1	4
SEQ ID NO: 3235	attgatttaacaaaagt	6823	6842	SEQ ID NO: 4564	actgactcagagagaata	11404	11423	1	4
SEQ ID NO: 3236	ctgcagcagcttaagagac	6914	6933	SEQ ID NO: 4565	gtctctcagtgaaagctcag	10699	10718	1	4
SEQ ID NO: 3237	aaacaacaactagaggct	6973	6992	SEQ ID NO: 4566	agctccacctcttacttt	10571	10590	1	4
SEQ ID NO: 3238	ctgagcatgctcaacacti	7059	7078	SEQ ID NO: 4567	aagtagctgagaaaactaa	7104	7123	1	4
SEQ ID NO: 3239	ctgagtagctgagaaaa	7100	7119	SEQ ID NO: 4568	tttccactatagatgcata	8421	8440	1	4
SEQ ID NO: 3240	ttgtagagtgtggccacc	7199	7218	SEQ ID NO: 4569	gttgagactctgctgctaa	7776	7795	1	4
SEQ ID NO: 3241	tgaggagagactatcagaa	7227	7246	SEQ ID NO: 4570	ttctcaatttgattttica	8526	8545	1	4
SEQ ID NO: 3242	gagactatcagagaactaa	7232	7251	SEQ ID NO: 4571	ttagccacagctctgtctc	10301	10320	1	4
SEQ ID NO: 3243	taattagtggaattattga	7293	7312	SEQ ID NO: 4572	ccaagaagcttaagtaatt	7320	7339	1	4
SEQ ID NO: 3244	gtcttaagaattattctt	7327	7346	SEQ ID NO: 4573	aaaacagagctctcagaagc	13209	13228	1	4
SEQ ID NO: 3245	taacaacttcctgacat	7365	7384	SEQ ID NO: 4574	gtgtctctacacaaagttaa	7254	7273	1	4
SEQ ID NO: 3246	aaattaaagcatttgatt	7394	7413	SEQ ID NO: 4575	aaactcttgacaggcattt	9723	9742	1	4
SEQ ID NO: 3247	gactcaatgggaaattca	7464	7483	SEQ ID NO: 4576	tgaaattcaatcacaagtc	9076	9095	1	4
SEQ ID NO: 3248	gaaattcaggctcgggaac	7475	7494	SEQ ID NO: 4577	gtctctcaatttgatttc	8525	8544	1	4
SEQ ID NO: 3249	actaccacaaagacgtaa	7492	7511	SEQ ID NO: 4578	ttcaggaactatgctclagt	10645	10664	1	4
SEQ ID NO: 3250	caaaaataacctaatcat	7578	7597	SEQ ID NO: 4579	tgatttccctgacactgg	10950	10969	1	4
SEQ ID NO: 3251	aaataaccttaactcaaa	7581	7600	SEQ ID NO: 4580	tgaaagtaagaagaaattt	10749	10768	1	4
SEQ ID NO: 3252	taagttcagcatctttg	7615	7634	SEQ ID NO: 4581	caaatctgagattcttaaa	9480	9499	1	4
SEQ ID NO: 3253	caggtttatgcacacttg	7739	7758	SEQ ID NO: 4582	caagggttcactgtctctc	7865	7884	1	4
SEQ ID NO: 3254	gttccactgttctgaaatc	7870	7889	SEQ ID NO: 4583	gattctcagatagggaac	8922	8941	1	4
SEQ ID NO: 3255	cactgttctcgaatcaag	7873	7892	SEQ ID NO: 4584	cttgaacacaaagtcagtc	6008	6027	1	4
SEQ ID NO: 3256	actgttctcgaatcaaga	7874	7893	SEQ ID NO: 4585	cttgaacacaaagtcagtc	6007	6026	1	4
SEQ ID NO: 3257	gctgtccttgaaagtcagt	7909	7928	SEQ ID NO: 4586	actgttactcaggaaggc	12580	12599	1	4
SEQ ID NO: 3258	aaacagatttgaggattcc	7980	7999	SEQ ID NO: 4587	ggaagctctcacaagattta	13222	13241	1	4
SEQ ID NO: 3259	gttttccaccaggaattt	8050	8069	SEQ ID NO: 4588	aaattttctcgtcggaac	9418	9437	1	4
SEQ ID NO: 3260	cagaaccattgaccagat	8136	8155	SEQ ID NO: 4589	atctgcagacaaagctgta	12438	12457	1	4
SEQ ID NO: 3261	tagcgagataccctgtcc	8226	8245	SEQ ID NO: 4590	ggcagctctctgctgtcta	12301	12320	1	4
SEQ ID NO: 3262	ctctaattgatttcaagt	8299	8318	SEQ ID NO: 4591	aactgttgactcaggaagg	12579	12598	1	4
SEQ ID NO: 3263	acataccagaattccagct	8328	8347	SEQ ID NO: 4592	agctgccagtcctctcatgt	10026	10045	1	4
SEQ ID NO: 3264	aatgctgacatagggaaag	8438	8457	SEQ ID NO: 4593	catatacttgcctcatcat	10005	10024	1	4
SEQ ID NO: 3265	atctgctgacatggaatgg	8439	8458	SEQ ID NO: 4594	ccatttgagatacaggcat	9245	9264	1	4
SEQ ID NO: 3266	aaccacctcagcaaacgta	8458	8477	SEQ ID NO: 4595	ctgtgttttccattaggtgt	9291	9310	1	4
SEQ ID NO: 3267	agcagtatcgagcttcc	8476	8495	SEQ ID NO: 4596	ggaagtgccctgaattgct	10972	10991	1	4
SEQ ID NO: 3268	tgcaaaccttcaaacct	8551	8570	SEQ ID NO: 4597	agggaaagagaagattgca	13501	13520	1	4
SEQ ID NO: 3269	ggagagctcagtgagttct	8592	8611	SEQ ID NO: 4598	ggaagactatcatcatct	13788	13807	1	4
SEQ ID NO: 3270	tttttgaaattccattga	8652	8671	SEQ ID NO: 4599	tcagtgaaattttatcaaaa	13194	13213	1	4
SEQ ID NO: 3271	aatgagtgattgtcaaga	8729	8748	SEQ ID NO: 4600	cttttcagccagccattt	9231	9250	1	4
SEQ ID NO: 3272	gtcagaataaacaatcagc	8741	8760	SEQ ID NO: 4601	gtcgtacttttaaaactcag	4819	4838	1	4
SEQ ID NO: 3273	tcacaataatcaactccc	8787	8806	SEQ ID NO: 4602	ggagatttccaaagctgtga	11772	11791	1	4
SEQ ID NO: 3274	tgaaacttcccaactcgg	8795	8814	SEQ ID NO: 4603	ccagtttctcaggagactcaa	12603	12622	1	4

SEQ ID NO: 3275	acatcccaaacctggactt	8799	8818	SEQ ID NO: 4604	aagtcgattccagcagtg	9090	9109	4
SEQ ID NO: 3276	actctctagtcagctgla	8814	8833	SEQ ID NO: 4605	ctcagatggaaaagaagt	11010	11029	4
SEQ ID NO: 3277	tgatcacaacatagtttc	8844	8863	SEQ ID NO: 4606	gaaagctccabattggtca	12817	12836	4
SEQ ID NO: 3278	agaaggacccctcacttc	8968	8987	SEQ ID NO: 4607	ggaagaaagagcagctct	12292	12311	4
SEQ ID NO: 3279	tggacgttccataaagat	8988	9007	SEQ ID NO: 4608	atctaagtgcagtagccaa	11634	11653	4
SEQ ID NO: 3280	actgttccataagatcaat	8992	9011	SEQ ID NO: 4609	atgataaaccatcacag	13891	13910	4
SEQ ID NO: 3281	ctgtccaataagatcaata	8993	9012	SEQ ID NO: 4610	atagataaaaaccatcacg	13890	13909	4
SEQ ID NO: 3282	ctttatgaatctggctcc	9041	9060	SEQ ID NO: 4611	gggaatctgatgaggaaac	12255	12274	4
SEQ ID NO: 3283	atgaaatctggctccctcaa	9045	9064	SEQ ID NO: 4612	atgagttgccaccatcat	11667	11686	4
SEQ ID NO: 3284	ctcaacttctaaactgt	9059	9078	SEQ ID NO: 4613	caagatcgcagactgtgag	11653	11672	4
SEQ ID NO: 3285	ctaaagctgctgactgtt	9129	9148	SEQ ID NO: 4614	aacgaacaaactgacttag	9749	9768	4
SEQ ID NO: 3286	aaggctgctgactgtttg	9132	9151	SEQ ID NO: 4615	ccaagaaaagccacacctt	11077	11096	4
SEQ ID NO: 3287	atccacacaacaaagagg	9262	9281	SEQ ID NO: 4616	ccctacagactttgaggat	7977	7996	4
SEQ ID NO: 3288	ggaatttgaaagtgttt	9279	9298	SEQ ID NO: 4617	aaacaaacacagccatcc	9655	9674	4
SEQ ID NO: 3289	aataactgtgactgtttc	9332	9351	SEQ ID NO: 4618	gaaactgttttctatt	12836	12855	4
SEQ ID NO: 3290	gaacaaagagcaactatt	9432	9451	SEQ ID NO: 4619	ataaacctgcaagattttc	13608	13627	4
SEQ ID NO: 3291	ttcttgaaacgcaacaagc	9599	9618	SEQ ID NO: 4620	gctttcaatgacaacagaa	11065	11084	4
SEQ ID NO: 3292	taagaaaaaacaacacag	9648	9667	SEQ ID NO: 4621	ctgtgctgttgagtttat	9690	9709	4
SEQ ID NO: 3293	aaacaaacacagcatttc	9654	9673	SEQ ID NO: 4622	gaatttgaaagtgtttt	9280	9299	4
SEQ ID NO: 3294	gcatctccatcacaaatct	9667	9686	SEQ ID NO: 4623	aggaagctgacctgaatgc	10971	10990	4
SEQ ID NO: 3295	ttgaaaaaacaacaaaca	9740	9759	SEQ ID NO: 4624	tttgaaagattatcaaca	12933	12952	4
SEQ ID NO: 3296	caatgacttagatttttc	9757	9776	SEQ ID NO: 4625	gacacaaaaaggggattg	10279	10298	4
SEQ ID NO: 3297	caaatgctgaaaaatctag	9817	9836	SEQ ID NO: 4626	ctgaaactctacatttg	11438	11457	4
SEQ ID NO: 3298	ctgggatacactgttccag	9863	9882	SEQ ID NO: 4627	ctggactctctagtcagg	8810	8829	4
SEQ ID NO: 3299	gttgaagtgtctcatcca	9890	9909	SEQ ID NO: 4628	tgaaatctggctccctaac	9046	9065	4
SEQ ID NO: 3300	ttctccatctaggttct	9964	9983	SEQ ID NO: 4629	agaatccagatacaagaaa	6893	6912	4
SEQ ID NO: 3301	ttctccatctaggttctg	9965	9984	SEQ ID NO: 4630	cagaatccagatacaagaa	6892	6911	4
SEQ ID NO: 3302	ctaatgagctgcagctcc	10019	10038	SEQ ID NO: 4631	ggacagtgaaattatgca	13305	13324	4
SEQ ID NO: 3303	tgctgaacttttaaacag	10177	10196	SEQ ID NO: 4632	ctggatgtaaccaccaga	11186	11205	4
SEQ ID NO: 3304	dcctttctcatcttcat	10214	10233	SEQ ID NO: 4633	atgaagactgtctcaggag	13772	13791	4
SEQ ID NO: 3305	gtctattgatgcacgtcag	10234	10253	SEQ ID NO: 4634	ctgcgtcactgaagaagaca	12080	12099	4
SEQ ID NO: 3306	gaatgcactgcagctacaaa	10240	10259	SEQ ID NO: 4635	tttgatgtgccaccatcca	11666	11685	4
SEQ ID NO: 3307	agctctgtctctgcagcaac	10308	10328	SEQ ID NO: 4636	gtttaccacaaagcttagct	10547	10566	4
SEQ ID NO: 3308	agccgaatccaattttg	10408	10427	SEQ ID NO: 4637	caagctgcacacagggct	13971	13990	4
SEQ ID NO: 3309	tggaaatgattcaataagc	10424	10443	SEQ ID NO: 4638	gcttcaggaagcttctcaa	13216	13235	4
SEQ ID NO: 3310	aaactactgtctctcttc	10469	10488	SEQ ID NO: 4639	aggaagccacagccattt	12591	12610	4
SEQ ID NO: 3311	tactttacttgatgtcat	10583	10602	SEQ ID NO: 4640	atgattatgtcaacaagta	12363	12382	4
SEQ ID NO: 3312	caagtctcagtcagctag	10918	10937	SEQ ID NO: 4641	ctgacactctagagcaatga	5001	5020	4
SEQ ID NO: 3313	atgcagatcagccagttc	10926	10945	SEQ ID NO: 4642	gaactcagaaagatggcat	14002	14021	4
SEQ ID NO: 3314	tgaaatgctaacataagaa	10983	11002	SEQ ID NO: 4643	ttctcaattttgatttcca	8526	8545	4
SEQ ID NO: 3315	agaagatcagatggaaaaa	11004	11023	SEQ ID NO: 4644	ttttctaaatggaaactct	12173	12192	4
SEQ ID NO: 3316	ggctattctatctcatcc	11264	11283	SEQ ID NO: 4645	ggatctaaatgctagatgc	11632	11651	4
SEQ ID NO: 3317	aaagtgttgctgataaat	11288	11307	SEQ ID NO: 4646	atgacttaaacattcttt	9489	9508	4
SEQ ID NO: 3318	agttttgtctgataaattc	11290	11309	SEQ ID NO: 4647	gaatctgctccctcaact	9047	9066	4
SEQ ID NO: 3319	ctgggtgctgaacaaatga	11316	11335	SEQ ID NO: 4648	ctcttctggctcttccag	11035	11054	4
SEQ ID NO: 3320	cagagaatacacaattat	11413	11432	SEQ ID NO: 4649	atagactgactctctctg	8873	8892	4
SEQ ID NO: 3321	gaggtaaaattccctgaag	11480	11499	SEQ ID NO: 4650	ctctgtgctgcttaacct	12306	12325	4

SEQ ID NO:	3322	ctttttgagataacogtg	11545	11564	SEQ ID NO:	4651	acaggagttactgaaaaag	13723	13742	4
SEQ ID NO:	3323	ctgggaattgcaattcct	11735	11754	SEQ ID NO:	4652	aaggatctccacccagc	12102	12121	4
SEQ ID NO:	3324	gtgataatgacccattg	11795	11814	SEQ ID NO:	4653	tccaagatgagcaaac	13104	13123	4
SEQ ID NO:	3325	attccacatcgacgtcaac	11859	11878	SEQ ID NO:	4654	gttgagaagccccaagaat	6254	6273	4
SEQ ID NO:	3326	gaagaagatgacaaattt	11992	12011	SEQ ID NO:	4655	aaattctctttcttcca	9220	9239	4
SEQ ID NO:	3327	ctcaaaagccagatcgcca	12050	12069	SEQ ID NO:	4656	gtgaagtcagcatctgt	12659	12688	4
SEQ ID NO:	3328	gtggcgatgataagtg	12143	12162	SEQ ID NO:	4657	catctctaaacccctcac	8071	8090	4
SEQ ID NO:	3329	aaatggaactctactaca	12179	12198	SEQ ID NO:	4658	gtaccataagccatattt	10088	10107	4
SEQ ID NO:	3330	aaaaactcaactattcaa	12219	12238	SEQ ID NO:	4659	gtgatgtagagctgtttt	6993	7012	4
SEQ ID NO:	3331	ctgagaagaaatctgaga	12428	12447	SEQ ID NO:	4660	ctgcagagaataatcag	13447	13466	4
SEQ ID NO:	3332	caaatctgagtggtttta	12447	12466	SEQ ID NO:	4661	aaatggagctttatgt	14086	14105	4
SEQ ID NO:	3333	caatctgagtggtttat	12448	12467	SEQ ID NO:	4662	ataaattggatctttatg	14085	14104	4
SEQ ID NO:	3334	taggcaaatgtatgatat	12477	12496	SEQ ID NO:	4663	ctatgttcagtgccctcaa	13392	13411	4
SEQ ID NO:	3335	ataaactaatagatgtaat	12897	12916	SEQ ID NO:	4664	atactatgtaaaaatttat	13641	13660	4
SEQ ID NO:	3336	ccaactaatagaagataac	13039	13058	SEQ ID NO:	4665	gtattttgtctaacttgg	14052	14071	4
SEQ ID NO:	3337	ttaattatccaagatga	13095	13114	SEQ ID NO:	4666	catcctctaatttttaa	13800	13819	4
SEQ ID NO:	3338	ttaaatgttgaagaagaa	13151	13170	SEQ ID NO:	4667	tttcatgtgaagaataaa	7032	7051	4
SEQ ID NO:	3339	aaatccaatgaattttat	13190	13209	SEQ ID NO:	4668	gaataccaatgctgaactt	10168	10187	4
SEQ ID NO:	3340	tgagaagaaatagatcgag	13326	13345	SEQ ID NO:	4669	ctgagagaagtgcttcaa	12407	12426	4
SEQ ID NO:	3341	acttctactctgaaatat	13377	13396	SEQ ID NO:	4670	ataatctggaacctgaagt	10737	10756	4
SEQ ID NO:	3342	ccacagaataatctgagaat	13451	13470	SEQ ID NO:	4671	tattccctgaagtgtgagtg	11488	11507	4
SEQ ID NO:	3343	ccattgcgacgaagaatat	13560	13579	SEQ ID NO:	4672	attttttctctgccaag	10103	10122	4
SEQ ID NO:	3344	tataaactgcaagattttt	13607	13626	SEQ ID NO:	4673	aaaattcaaacgtccata	13873	13892	4
SEQ ID NO:	3345	ctgtattactatgaaaaat	13637	13656	SEQ ID NO:	4674	tattgttaagaataacaga	6436	6455	4
SEQ ID NO:	3346	ggagttactgaaaaaagctg	13726	13745	SEQ ID NO:	4675	cagcatgctcgatttctcc	9952	9971	4
SEQ ID NO:	3347	gaagctgtctcaggaga	13773	13792	SEQ ID NO:	4676	tccctcttctacttcca	10213	10232	4
SEQ ID NO:	3348	gaactggaactgcacaaa	13955	13974	SEQ ID NO:	4677	tgttgtagacaagggttca	7856	7875	4
SEQ ID NO:	3349	gtctaaactgtggggagtg	14058	14077	SEQ ID NO:	4678	ctctctacagtggtggcaa	4230	4249	4
SEQ ID NO:	3350	gattcgaaatatacaattca	4412	4431	SEQ ID NO:	4679	gaaaacgacaagaacatc	9603	9622	3
SEQ ID NO:	3351	atttgtttgtaagaagtg	4551	4570	SEQ ID NO:	4680	acttttctaacttgaaat	9063	9082	3
SEQ ID NO:	3352	tctgtgtgtctgctgcg	33	52	SEQ ID NO:	4681	tcagccagccatttgaga	9236	9255	3
SEQ ID NO:	3353	ctgtgagctgcgcgccagc	47	66	SEQ ID NO:	4682	ctgtgtgattgaaccacagc	11185	11204	3
SEQ ID NO:	3354	ctgtgtcttccaaaagatg	227	246	SEQ ID NO:	4683	catcagaacccattgaacag	8134	8153	3
SEQ ID NO:	3355	ctgagagttccagtgagtg	291	310	SEQ ID NO:	4684	actcaalgtggaattcag	7465	7484	3
SEQ ID NO:	3356	cagtgacccctgaagaagtg	404	423	SEQ ID NO:	4685	ctccactctcttggagtg	8977	8996	3
SEQ ID NO:	3357	ctctgaggagttgtctgca	472	491	SEQ ID NO:	4686	gtcaaacctgactcactag	11399	11418	3
SEQ ID NO:	3358	acatcaagaggggcatcat	582	601	SEQ ID NO:	4687	atgacgttcttgagcatgt	7050	7069	3
SEQ ID NO:	3359	ctgatcagcagccagct	830	849	SEQ ID NO:	4688	actggaacttctctagtcag	8809	8828	3
SEQ ID NO:	3360	ggagcgtgaagagaagcat	865	884	SEQ ID NO:	4689	atgctgctaagcttccatgctc	11354	11373	3
SEQ ID NO:	3361	agctgtttgaagactctc	1087	1106	SEQ ID NO:	4690	gagaagctgtcttcaaaagt	12411	12430	3
SEQ ID NO:	3362	tgaaaaaactaacactctc	1113	1132	SEQ ID NO:	4691	tgagatcaacacacttctca	13112	13131	3
SEQ ID NO:	3363	ctgagctgagcagctcag	1176	1195	SEQ ID NO:	4692	ctgaattactgcaactcag	3035	3054	3
SEQ ID NO:	3364	tgaagacgtgtgcatgcaca	1311	1330	SEQ ID NO:	4693	gtgtagaccaagggttca	7856	7875	3
SEQ ID NO:	3365	ctctgtatgctgctgacca	1440	1459	SEQ ID NO:	4694	tggcactgtttggagaagg	9138	9157	3
SEQ ID NO:	3366	aggagctgctgagcaattgc	1500	1519	SEQ ID NO:	4695	gcaagtcagccagttact	10928	10947	3
SEQ ID NO:	3367	attttgattctgcggctcat	1575	1594	SEQ ID NO:	4696	atgaaacaaatgacaaat	7428	7447	3
SEQ ID NO:	3368	tccagaactcaagttctca	1627	1646	SEQ ID NO:	4697	tgaatacaatgtctctgga	5520	5539	3

SEQ ID NO: 3369	gggttccttcagacttc	1744	1763	SEQ ID NO: 4698	gaataccaagctcaaac	10455	10474	2	3
SEQ ID NO: 3370	gttgatgaggagtcctca	1810	1829	SEQ ID NO: 4699	gcaaaagctgcaatcaac	13734	13753	2	3
SEQ ID NO: 3371	ccaagatctgaaaaagtt	1941	1960	SEQ ID NO: 4700	aaactgtcttccaaatgga	3552	3571	2	3
SEQ ID NO: 3372	agttagtgaaagaagttct	1956	1975	SEQ ID NO: 4701	agacttcataatcccaact	8275	8294	2	3
SEQ ID NO: 3373	gaagggaatctattattg	2084	2103	SEQ ID NO: 4702	caaacactactgtctctc	10467	10486	2	3
SEQ ID NO: 3374	gaagactcttttgggaag	2221	2240	SEQ ID NO: 4703	cttcacatacagaattcc	8324	8343	2	3
SEQ ID NO: 3375	ggaataatgctcactgtt	2374	2393	SEQ ID NO: 4704	caacaacacagacttcca	9656	9675	2	3
SEQ ID NO: 3376	gatttgaatccaagaag	2408	2427	SEQ ID NO: 4705	ctcatgtccctagaatac	10037	10056	2	3
SEQ ID NO: 3377	ccaagaagaagctccggaag	2417	2436	SEQ ID NO: 4706	ctcagctgtcttcttggga	4951	4970	2	3
SEQ ID NO: 3378	aggaaggctcaagaagtg	2570	2589	SEQ ID NO: 4707	catfagagctgcagctct	10020	10039	2	3
SEQ ID NO: 3379	gagaatgactttttctca	2583	2602	SEQ ID NO: 4708	gaagatgacagactttct	12160	12179	2	3
SEQ ID NO: 3380	ttgtgacaaatatgggca	2765	2784	SEQ ID NO: 4709	gcccagtttgaanaacaaa	11815	11834	2	3
SEQ ID NO: 3381	ctgaggctcccatgacatt	3252	3271	SEQ ID NO: 4710	aalgctagctcttgttcaag	10903	10922	2	3
SEQ ID NO: 3382	gtagatccaaaaaaatga	3668	3687	SEQ ID NO: 4711	ttcattgcctccaactcac	11450	11469	2	3
SEQ ID NO: 3383	aaatgacttccaatttccc	3681	3700	SEQ ID NO: 4712	gggaactgtgaaagatt	12927	12946	2	3
SEQ ID NO: 3384	atgacttccaatttccgtg	3683	3702	SEQ ID NO: 4713	caggagacttactatcat	13785	13804	2	3
SEQ ID NO: 3385	atctgcacatctgagatt	4104	4123	SEQ ID NO: 4714	aacctctccatcgaagaat	9547	9566	2	3
SEQ ID NO: 3386	attttgttgcgaagaagt	4551	4570	SEQ ID NO: 4715	acttcogtttaccagaat	8247	8266	2	3
SEQ ID NO: 3387	gcagagcttggcctctctg	5135	5154	SEQ ID NO: 4716	cagagcttctgcacactgc	13518	13537	2	3
SEQ ID NO: 3388	atagtctgaatgaatt	5353	5372	SEQ ID NO: 4717	aaatccaactgctctat	13874	13893	2	3
SEQ ID NO: 3389	ccaacacttgacaacatt	5420	5439	SEQ ID NO: 4718	aaatacttccacaatattga	8780	8799	2	3
SEQ ID NO: 3390	cagtgacctgaaatacaat	5512	5531	SEQ ID NO: 4719	attgaacatccccaaactg	8794	8813	2	3
SEQ ID NO: 3391	tacaatggcgaatgggaaa	5848	5867	SEQ ID NO: 4720	ttcaactgccttttgttga	11229	11248	2	3
SEQ ID NO: 3392	ctttgtlaaagtatgataa	6285	6304	SEQ ID NO: 4721	ttatgtctgaatccaaaag	13656	13675	2	3
SEQ ID NO: 3393	ttgtaaagtatgataaaaa	6288	6307	SEQ ID NO: 4722	ttttcaagcaaatgacacaa	8539	8558	2	3
SEQ ID NO: 3394	lccattaacctccatttt	6320	6339	SEQ ID NO: 4723	aaaaagaaaatttgtctgaa	10756	10775	2	3
SEQ ID NO: 3395	gattatctgaattcatcca	6488	6507	SEQ ID NO: 4724	tgagtaglacacaacaatc	7162	7181	2	3
SEQ ID NO: 3396	aatttgggagagacaaagt	6506	6525	SEQ ID NO: 4725	aaactaaatgatctaatt	11324	11343	2	3
SEQ ID NO: 3397	atttgaataatgctatgac	6696	6715	SEQ ID NO: 4726	gcaattctgcacagaat	13441	13460	2	3
SEQ ID NO: 3398	tgagcatgtcaaacacttt	7060	7079	SEQ ID NO: 4727	aaagccattcagctctcca	12971	12990	2	3
SEQ ID NO: 3399	ttgaagatgttacaacatt	7356	7375	SEQ ID NO: 4728	aaatccatgatgaagtcaa	12660	12679	2	3
SEQ ID NO: 3400	acttgcacttcaacttct	7753	7772	SEQ ID NO: 4729	gaagatattttgttccaagt	13276	13295	2	3
SEQ ID NO: 3401	ttttccacaccagaatt	8050	8069	SEQ ID NO: 4730	aaatcttgatttctaatac	9481	9500	2	3
SEQ ID NO: 3402	ataagtacaacaaaattt	9405	9424	SEQ ID NO: 4731	aaataaatggagctttat	14083	14102	2	3
SEQ ID NO: 3403	cggtgacctgcgggtctgag	8	27	SEQ ID NO: 4732	ctcagtttaactcgttccog	11571	11590	2	3
SEQ ID NO: 3404	agtgcccttctcggttgt	25	44	SEQ ID NO: 4733	agcatctgattgactact	12678	12697	2	3
SEQ ID NO: 3405	gctgaggagccgccccagc	47	66	SEQ ID NO: 4734	gctgattgaggtgtccagc	1225	1244	2	3
SEQ ID NO: 3406	agaggagccgcgcagccag	50	69	SEQ ID NO: 4735	cttgatcacagagctctcc	3752	3771	2	3
SEQ ID NO: 3407	ggccgcgagggccgagcc	72	91	SEQ ID NO: 4736	ggcctgatcccgagccc	1363	1382	2	3
SEQ ID NO: 3408	ccaggccgagcccgagcc	89	108	SEQ ID NO: 4737	ctcccgaggccaagcgttg	2682	2701	2	3
SEQ ID NO: 3409	ggagccgcgccacccgcgc	104	123	SEQ ID NO: 4738	gctgttttgcagactctcc	1088	1107	2	3
SEQ ID NO: 3410	gaagaggaaatgtctgaaa	200	219	SEQ ID NO: 4739	ttcaagattcttgacattc	8309	8328	2	3
SEQ ID NO: 3411	caaaagtgcaccagcgaat	237	256	SEQ ID NO: 4740	aaactctatttgggattttg	7085	7104	2	3
SEQ ID NO: 3412	attcaagcactccoggaag	253	272	SEQ ID NO: 4741	cttcacatttcaaggaat	10067	10086	2	3
SEQ ID NO: 3413	cttcagtgatcttccttgg	297	316	SEQ ID NO: 4742	ccagcaagctcgtgagaac	8610	8629	2	3
SEQ ID NO: 3414	gactgctgattcaagaagt	316	335	SEQ ID NO: 4743	acttgaagaaagatagtc	13324	13343	2	3
SEQ ID NO: 3415	gtgccaccaggatcaactg	333	352	SEQ ID NO: 4744	cagtgaagctgcaggcgac	10704	10723	2	3

SEQ ID NO: 3416	gatcaactgcaggttgag	343	362	SEQ ID NO: 4745	ctcaactccaactcgtac	4748	4767	3
SEQ ID NO: 3417	actgcaggttgagctgga	348	367	SEQ ID NO: 4746	ccactacacactccactg	1289	1308	3
SEQ ID NO: 3418	ccagctctgcagcttcac	373	392	SEQ ID NO: 4747	gattgtgtccactactcgtg	1343	1362	3
SEQ ID NO: 3419	agcttactctgaagacaa	383	402	SEQ ID NO: 4748	gggtcgtggagatgagct	5112	5131	3
SEQ ID NO: 3420	cttactctgaagaccagc	385	404	SEQ ID NO: 4749	gctggagtaaacctggaag	2696	2715	3
SEQ ID NO: 3421	ccagccagtgccactcga	399	418	SEQ ID NO: 4750	ttcaagatgactcactgag	1539	1558	3
SEQ ID NO: 3422	cagtgccactgaaagagg	404	423	SEQ ID NO: 4751	gctcacagactcactcactg	5230	5249	3
SEQ ID NO: 3423	ggcttcaacccctgagggc	427	446	SEQ ID NO: 4752	gcccactgggtgcctgccca	3533	3552	3
SEQ ID NO: 3424	cttcaacccctgagggcaca	430	449	SEQ ID NO: 4753	tttgagccaacttggaag	2207	2226	3
SEQ ID NO: 3425	ttcaacccctgagggcacaag	431	450	SEQ ID NO: 4754	ctttgacaggctatttgag	9727	9746	3
SEQ ID NO: 3426	cttgctgaagaaacccaag	451	470	SEQ ID NO: 4755	cttgaattcaatcacacaag	9074	9093	3
SEQ ID NO: 3427	gctgaagaaacccaagaa	453	472	SEQ ID NO: 4756	ttctgctgcttactcagca	5647	5666	3
SEQ ID NO: 3428	ttgctgcagccatgtccag	483	502	SEQ ID NO: 4757	ctggctcatttgcaagcaa	3004	3023	3
SEQ ID NO: 3429	gctgcagccatgtccagg	484	503	SEQ ID NO: 4758	ctctgctcattgttcgaagca	3003	3022	3
SEQ ID NO: 3430	ggccatgtccaggttagag	490	509	SEQ ID NO: 4759	ctcaactcctcaggctgct	1293	1312	3
SEQ ID NO: 3431	ggctcagctggccatctcc	507	526	SEQ ID NO: 4760	ggaaactaccacaaaagact	7489	7508	3
SEQ ID NO: 3432	ggaagggaagcaggttttc	526	545	SEQ ID NO: 4761	gaaattctcattattctt	13821	13840	3
SEQ ID NO: 3433	aagggaagcaggttttct	528	547	SEQ ID NO: 4762	aggacacccaataaacctt	7572	7591	3
SEQ ID NO: 3434	ggaagaatgaacctactta	555	574	SEQ ID NO: 4763	taagaactttgccactctt	4852	4871	3
SEQ ID NO: 3435	atctcgaatcaacgaaggg	575	594	SEQ ID NO: 4764	ccctaacagattgaggat	7977	7996	3
SEQ ID NO: 3436	ctctgaacatcaagagggg	576	595	SEQ ID NO: 4765	ccctaacacagattgagga	7976	7995	3
SEQ ID NO: 3437	ctgaacatcaagaggggcca	578	597	SEQ ID NO: 4766	ggcctgctttgaggtcag	7908	7927	3
SEQ ID NO: 3438	aactcaagagggggcatca	581	600	SEQ ID NO: 4767	gtataaaaaaccaagattt	6298	6317	3
SEQ ID NO: 3439	acatcaagaggggcatcat	582	601	SEQ ID NO: 4768	attgataaaaaaccaagattg	6297	6316	3
SEQ ID NO: 3440	catcttctgcctctcggt	597	616	SEQ ID NO: 4769	accaccagtttgtatgta	7413	7432	3
SEQ ID NO: 3441	ttcccccagagacagaaga	615	634	SEQ ID NO: 4770	cttccacatttcaaggaa	10066	10085	3
SEQ ID NO: 3442	gaagaagccaagcagatgt	629	648	SEQ ID NO: 4771	acactctccactctcttc	8079	8098	3
SEQ ID NO: 3443	tggtttctggtacatcgtt	647	666	SEQ ID NO: 4772	acactcaataactctccaca	8775	8794	3
SEQ ID NO: 3444	tgatgtgaaactgtctccac	663	682	SEQ ID NO: 4773	gtggaggccaacacattaca	2928	2947	3
SEQ ID NO: 3445	aaactgtctcactacttt	670	689	SEQ ID NO: 4774	aaagaacacgcatttgttt	4540	4559	3
SEQ ID NO: 3446	actcacttttaccgtcaaga	680	699	SEQ ID NO: 4775	cttacttttctattgatt	10580	10599	3
SEQ ID NO: 3447	tttaccgttcaagcagagg	685	704	SEQ ID NO: 4776	ctccactcgtctgggaag	2491	2510	3
SEQ ID NO: 3448	ttaccgttcaagcagaggaa	687	706	SEQ ID NO: 4777	ttccataagcttgatgtaa	11177	11196	3
SEQ ID NO: 3449	acagaggaaggcattgtgg	698	717	SEQ ID NO: 4778	ccacaaagctcatctcgtg	5964	5983	3
SEQ ID NO: 3450	gcagggaaggcattgtggc	699	718	SEQ ID NO: 4779	gccagaagtgagactcctg	3515	3534	3
SEQ ID NO: 3451	gagggaaggcattgtggca	700	719	SEQ ID NO: 4780	gtccagcttccatgactct	2476	2495	3
SEQ ID NO: 3452	gggaaggcattgtggcaac	702	721	SEQ ID NO: 4781	gtgtcttaaggacttcc	13364	13383	3
SEQ ID NO: 3453	gaagggcattgtggcaaca	703	722	SEQ ID NO: 4782	gtgtgtagaggacttctc	1809	1828	3
SEQ ID NO: 3454	caggcatcagccactctgc	777	796	SEQ ID NO: 4783	gccaagtcttctcggcct	3019	3038	3
SEQ ID NO: 3455	aggcatcagccactctgct	778	797	SEQ ID NO: 4784	agcaagcttcttctggcct	3018	3037	3
SEQ ID NO: 3456	tcagccactgtctctcat	783	802	SEQ ID NO: 4785	atgaaagtcaagactctga	12668	12687	3
SEQ ID NO: 3457	gtcaactctgatcagcagc	823	842	SEQ ID NO: 4786	gctgactttaaattctgac	4819	4838	3
SEQ ID NO: 3458	ggagcctaagggaagcat	865	884	SEQ ID NO: 4787	gtgcactgtttcagtcct	9339	9358	3
SEQ ID NO: 3459	aaggagcaacactcttcc	902	921	SEQ ID NO: 4788	gggaatatcttagcatctct	13465	13484	3
SEQ ID NO: 3460	aggagcaacacactcttcc	903	922	SEQ ID NO: 4789	aggaatatcttagcatctct	13464	13483	3
SEQ ID NO: 3461	caacactcttctctcct	908	927	SEQ ID NO: 4790	aaaggctgactctgtgttg	4292	4311	3
SEQ ID NO: 3462	aaacactcttctctcctt	909	928	SEQ ID NO: 4791	aaagcaggcggcagactgtt	1075	1094	3

SEQ ID NO: 3463	acaagaataagtagggat	933	952	SEQ ID NO: 4792	atccatgatctacattgt	6794	6813	3
SEQ ID NO: 3464	caagaataagtagggatg	934	953	SEQ ID NO: 4793	catcaacttacaagccttg	1246	1265	3
SEQ ID NO: 3465	agcacaagatgacacagac	954	973	SEQ ID NO: 4794	gtctctctgttctatccta	4592	4611	3
SEQ ID NO: 3466	agcacaagtgacacagact	955	974	SEQ ID NO: 4795	agctctctgttctatct	4591	4610	3
SEQ ID NO: 3467	agcacaagtgacacagact	956	975	SEQ ID NO: 4796	aagtgtagtctcctgtgac	5099	5118	3
SEQ ID NO: 3468	actctgaagacacacccaa	978	997	SEQ ID NO: 4797	tttagaggatccatagtt	7987	8006	3
SEQ ID NO: 3469	ctctcttggtgaaggtac	1008	1027	SEQ ID NO: 4798	gtactctactctgtgcaagc	8372	8391	3
SEQ ID NO: 3470	cttggtgaaggtactaag	1012	1031	SEQ ID NO: 4799	cttatgggatttctcaaac	11167	11186	3
SEQ ID NO: 3471	tactaagaagatgggcttc	1024	1043	SEQ ID NO: 4800	ggaggtgatcfaaacagta	10337	10356	3
SEQ ID NO: 3472	tttgagagacacaaatcca	1046	1065	SEQ ID NO: 4801	tggaagtgctgacggcaaa	10380	10399	3
SEQ ID NO: 3473	agagacacaaatccacatc	1050	1069	SEQ ID NO: 4802	gatggalatgaccttctct	4876	4895	3
SEQ ID NO: 3474	gttggttttggaagatctct	1087	1106	SEQ ID NO: 4803	gagaaacatactggcgact	5880	5899	3
SEQ ID NO: 3475	tgaaaaaactaacaccatct	1113	1132	SEQ ID NO: 4804	ggagaaatcaatgcctcca	7112	7131	3
SEQ ID NO: 3476	aaaaaaactaacaccatctct	1114	1133	SEQ ID NO: 4805	agagccagctgcagacttcc	11052	11071	3
SEQ ID NO: 3477	cttgagcaaaatatccaga	1130	1149	SEQ ID NO: 4806	ctctgatggaaaactcaga	12280	12299	3
SEQ ID NO: 3478	ttctcttaataagcttggt	1156	1175	SEQ ID NO: 4807	aacctccactttttgaga	6326	6345	3
SEQ ID NO: 3479	ctgagctgagagggcctcag	1176	1195	SEQ ID NO: 4808	ctgatcccgagccctcag	1367	1386	3
SEQ ID NO: 3480	tgagcagtcacatctctc	1198	1217	SEQ ID NO: 4809	ggagaaatcaatgcctcca	7112	7131	3
SEQ ID NO: 3481	agcagtcacatctctctct	1200	1219	SEQ ID NO: 4810	aagaggcagctcttggtct	12297	12316	3
SEQ ID NO: 3482	ctctctgcccacagctgat	1212	1231	SEQ ID NO: 4811	tatacaagagagcccacagag	12946	12965	3
SEQ ID NO: 3483	cttgccacagctgattga	1215	1234	SEQ ID NO: 4812	caaaagttaattgggaaga	12279	12298	3
SEQ ID NO: 3484	ctggccacagctgattgag	1216	1235	SEQ ID NO: 4813	ctcaatttggatttcaag	8528	8547	3
SEQ ID NO: 3485	tgagggtgtccagcccatc	1231	1250	SEQ ID NO: 4814	gatggaacctctccctcca	4733	4752	3
SEQ ID NO: 3486	tcagtgaggacagcctcag	1267	1286	SEQ ID NO: 4815	ctgacatcttaggcactga	5001	5020	3
SEQ ID NO: 3487	acatccctcagctggctgaa	1296	1315	SEQ ID NO: 4816	ctcagaagctaagcaatgt	7239	7258	3
SEQ ID NO: 3488	gcacagcagctgcgagaga	1385	1404	SEQ ID NO: 4817	ctctgtaagacaacagctgc	12323	12342	3
SEQ ID NO: 3489	cagcagctgcgagagatct	1388	1407	SEQ ID NO: 4818	agataacattaaacagctgc	13051	13070	3
SEQ ID NO: 3490	ggagggatcagcgacgccc	1415	1434	SEQ ID NO: 4819	ggctcaacacagacatgc	5718	5737	3
SEQ ID NO: 3491	aagacaaacctacaggga	1478	1497	SEQ ID NO: 4820	ccagcaaaacctcttctt	3936	3955	3
SEQ ID NO: 3492	caggagctgctggacattc	1499	1518	SEQ ID NO: 4821	caatggagagctcaaacctg	4660	4679	3
SEQ ID NO: 3493	aggagctgctggacattc	1500	1519	SEQ ID NO: 4822	gcaagggttctactgtctct	7864	7883	3
SEQ ID NO: 3494	ctgctggacattgctaat	1505	1524	SEQ ID NO: 4823	aattgggaagagagggcag	12287	12306	3
SEQ ID NO: 3495	gattacaacctattgattc	1565	1584	SEQ ID NO: 4824	gaatatattgagaggaaac	6353	6372	3
SEQ ID NO: 3496	attgattctgcgggtcat	1575	1594	SEQ ID NO: 4825	atgaagtagaccacaat	7161	7180	3
SEQ ID NO: 3497	ctctcggtctatctgaaat	1582	1601	SEQ ID NO: 4826	atttgttaagaaaatcaga	6436	6455	3
SEQ ID NO: 3498	aacctggagcagtttaact	1609	1628	SEQ ID NO: 4827	atttctccatctataggt	9962	9981	3
SEQ ID NO: 3499	ggagcagtttaactccagaa	1615	1634	SEQ ID NO: 4828	ttctgaaatccaactctc	8400	8419	3
SEQ ID NO: 3500	actccagaactcaagctct	1625	1644	SEQ ID NO: 4829	aagatgcagactttgattg	11654	11673	3
SEQ ID NO: 3501	ccagaaactcaagctctca	1627	1646	SEQ ID NO: 4830	gaactcagagaagattgga	1920	1939	3
SEQ ID NO: 3502	aagtacaaagccatcactg	1663	1682	SEQ ID NO: 4831	cagtcagtagaagaaactt	4429	4448	3
SEQ ID NO: 3503	gcccactcagatgactccag	1672	1691	SEQ ID NO: 4832	ctggaaactctctccatgac	10883	10902	3
SEQ ID NO: 3504	ccactcagtagatccagca	1673	1692	SEQ ID NO: 4833	ctgaactcagaagattg	13999	14018	3
SEQ ID NO: 3505	atccagaagctgcatcc	1685	1704	SEQ ID NO: 4834	gatttctcaagcttgat	11173	11192	3
SEQ ID NO: 3506	cagaagagctgcatccag	1688	1707	SEQ ID NO: 4835	ctgaaatacaatgctctg	5518	5537	3
SEQ ID NO: 3507	acaaggaccagaggttct	1731	1750	SEQ ID NO: 4836	agaacagcattgttgtt	4542	4561	3
SEQ ID NO: 3508	aggaacaggaggttctct	1734	1753	SEQ ID NO: 4837	gaagctcaagcaatgctct	7242	7261	3
SEQ ID NO: 3509	accagagggttctcttca	1737	1756	SEQ ID NO: 4838	tgagggtgactctgtgt	4290	4309	3

SEQ ID NO: 3510	cttcagacttctctgat	1750	1769	SEQ ID NO: 4839	atcaggaagggtcaaaga	2567	2586	3
SEQ ID NO: 3511	ctcaagcttctctgatga	1752	1771	SEQ ID NO: 4840	ctattactctctggctga	11307	11326	3
SEQ ID NO: 3512	gtgatgaggagctctca	1810	1829	SEQ ID NO: 4841	tgatctggctcctcaac	9046	9065	3
SEQ ID NO: 3513	cttcacaggagcatataa	1824	1843	SEQ ID NO: 4842	taatcgaggagtgatga	7148	7167	3
SEQ ID NO: 3514	cttcacaggcagcatataac	1825	1844	SEQ ID NO: 4843	gttaatcgaggagtgatga	7147	7166	3
SEQ ID NO: 3515	gtccagatataacaaat	1831	1850	SEQ ID NO: 4844	taattgcatatgatgtcc	6589	6608	3
SEQ ID NO: 3516	gataatacaaaattgtcca	1836	1855	SEQ ID NO: 4845	tgaggttgtagcaaat	2760	2779	3
SEQ ID NO: 3517	gacaaattgtccaaattct	1842	1861	SEQ ID NO: 4846	agaagaacagctgtttgt	4542	4561	3
SEQ ID NO: 3518	ggacaggtgaagaacttgg	1877	1896	SEQ ID NO: 4847	caaatgacatgatgggtc	5334	5353	3
SEQ ID NO: 3519	gtgaagaacttggctggct	1883	1902	SEQ ID NO: 4848	aagcatctgatgactacc	12677	12696	3
SEQ ID NO: 3520	ggaacttggctgtccca	1887	1906	SEQ ID NO: 4849	ggggctgtcccagattct	8909	8928	3
SEQ ID NO: 3521	tttggctccocatattg	1892	1911	SEQ ID NO: 4850	caataagatcaatagcaa	8998	9017	3
SEQ ID NO: 3522	tgctccocatattgccaa	1896	1915	SEQ ID NO: 4851	ttggctcacatgaaggcca	7631	7650	3
SEQ ID NO: 3523	ctccatattgccaatatc	1900	1919	SEQ ID NO: 4852	gatatacatagggaggaa	12745	12764	3
SEQ ID NO: 3524	ctccatattgccaatatc	1901	1920	SEQ ID NO: 4853	agatcaaatgtaattggga	12278	12297	3
SEQ ID NO: 3525	ttgcaatatcttgcaactc	1908	1927	SEQ ID NO: 4854	gagtcctcgatccagcaa	9352	9371	3
SEQ ID NO: 3526	ttggatatccaagatcga	1934	1953	SEQ ID NO: 4855	tcagtataagaacaacaa	9400	9419	3
SEQ ID NO: 3527	ccaagatctgaaaaggt	1941	1960	SEQ ID NO: 4856	aaactccaactgtcatgga	1986	2005	3
SEQ ID NO: 3528	ctgaaaaagttagtgaag	1948	1968	SEQ ID NO: 4857	cttgaagtacgtctcag	7915	7934	3
SEQ ID NO: 3529	agttagtgaagaagctgt	1956	1975	SEQ ID NO: 4858	agaatctcaactccaact	1978	1997	3
SEQ ID NO: 3530	aatctcaactctgaactgt	1980	1999	SEQ ID NO: 4859	acaggggtctctatgatt	12350	12369	3
SEQ ID NO: 3531	gtcatgactctcagaaat	1997	2016	SEQ ID NO: 4860	attgaaagaaataatgac	7036	7055	3
SEQ ID NO: 3532	ctcaactctcaaatctgt	2029	2048	SEQ ID NO: 4861	taacattgaggctattga	6978	6997	3
SEQ ID NO: 3533	aaactctcaaatctgttct	2031	2050	SEQ ID NO: 4862	gaaaaaagggttagaatt	10284	10303	3
SEQ ID NO: 3534	aaatagaagggaatcttat	2079	2098	SEQ ID NO: 4863	ataagcaaatctgtaatt	5457	5476	3
SEQ ID NO: 3535	gaaagggaatcttatatt	2083	2102	SEQ ID NO: 4864	aaatgcactgctgcttct	4900	4919	3
SEQ ID NO: 3536	gaagggaatcttatattg	2084	2103	SEQ ID NO: 4865	caaaaacatttcaactc	5287	5306	3
SEQ ID NO: 3537	gatccaaataactactct	2101	2120	SEQ ID NO: 4866	gaagggaagaagaaaaatca	3461	3480	3
SEQ ID NO: 3538	tgatttgctcagctgac	2158	2177	SEQ ID NO: 4867	gtcagccagttctctcca	10932	10951	3
SEQ ID NO: 3539	ttgtctcagctgacctca	2162	2181	SEQ ID NO: 4868	tgaggaaactcagatcaaa	12265	12284	3
SEQ ID NO: 3540	cttgaagggaaggctt	2191	2210	SEQ ID NO: 4869	aaagcatgtgtagacga	7850	7869	3
SEQ ID NO: 3541	tggaaggaaaaggcttga	2193	2212	SEQ ID NO: 4870	ctcaagtcgttggaatcca	4086	4105	3
SEQ ID NO: 3542	gtggttgagccaactgtg	2204	2223	SEQ ID NO: 4871	ccaagaggtatttaagacc	12958	12977	3
SEQ ID NO: 3543	tgagccaactgtgaagct	2209	2228	SEQ ID NO: 4872	agctttctgccaactgtcca	13521	13540	3
SEQ ID NO: 3544	gagccaactgtgaagctc	2210	2229	SEQ ID NO: 4873	gagctttctccaactgtct	13520	13539	3
SEQ ID NO: 3545	gaacttgaagctctttt	2215	2234	SEQ ID NO: 4874	aaaaagaacacagattgtt	4539	4558	3
SEQ ID NO: 3546	tggaagctcttttgggaa	2220	2239	SEQ ID NO: 4875	cttcgcacgtgtgggtcca	3785	3804	3
SEQ ID NO: 3547	ctctcttttgggaagcagg	2226	2245	SEQ ID NO: 4876	cttactgactgttgcagag	7798	7817	3
SEQ ID NO: 3548	ttttgggaagcaggatt	2229	2248	SEQ ID NO: 4877	taactgataaaataaaaa	6730	6749	3
SEQ ID NO: 3549	tttcacagcagtgtaaa	2247	2266	SEQ ID NO: 4878	tttgataaactgtaaaaa	6723	6742	3
SEQ ID NO: 3550	tttgctataccaagaatga	2331	2350	SEQ ID NO: 4879	catgtctccggagagccaa	2676	2695	3
SEQ ID NO: 3551	ttatacaaaagtataaaca	2337	2356	SEQ ID NO: 4880	gttctgtttgtaaatgat	6280	6299	3
SEQ ID NO: 3552	gagcaggatattgtaaatg	2357	2376	SEQ ID NO: 4881	catctcagctctgggtc	4262	4281	3
SEQ ID NO: 3553	tttggttaattggaataatgc	2366	2385	SEQ ID NO: 4882	gcagtcgactgttctcat	9954	9973	3
SEQ ID NO: 3554	ttgttaattggaataatgct	2367	2386	SEQ ID NO: 4883	agcagcagtagaanaacca	10809	10828	3
SEQ ID NO: 3555	aaattggaataatgctcag	2370	2389	SEQ ID NO: 4884	ctgaaagagatgaattttaa	13067	13086	3
SEQ ID NO: 3556	tggaataatgctcagttgt	2374	2393	SEQ ID NO: 4885	gaacagattgaggattcca	7981	8000	3

SEQ ID NO:	3557	icaglttgagaagctgat	2385	2404	SEQ ID NO:	4886	atcacaaclctocactga	9542	9561	3
SEQ ID NO:	3558	caglttgagaagctgatt	2386	2405	SEQ ID NO:	4887	aatcacaaclctocactg	9541	9560	3
SEQ ID NO:	3559	aglttgagaagctgatta	2387	2406	SEQ ID NO:	4888	atcacaaclctocactga	9540	9559	3
SEQ ID NO:	3560	ggttaagattggaalcc	2401	2420	SEQ ID NO:	4889	ggatactaagttaccaatc	6874	6893	3
SEQ ID NO:	3561	gattgaaatccaagaag	2408	2427	SEQ ID NO:	4890	cttcogttaccagaatc	8248	8267	3
SEQ ID NO:	3562	glttgaaatccaagaag	2409	2428	SEQ ID NO:	4891	acttcogttaccagaat	8247	8266	3
SEQ ID NO:	3563	atcacaagaagtcoggaa	2416	2435	SEQ ID NO:	4892	cttcaattccctgagat	3688	3707	3
SEQ ID NO:	3564	ccaagaagtcogggaag	2417	2436	SEQ ID NO:	4893	cttcaattccctgagga	3687	3706	3
SEQ ID NO:	3565	agagctacctccgcctct	2438	2457	SEQ ID NO:	4894	agattaatccgcctgct	8571	8590	3
SEQ ID NO:	3566	gagctacctccgcctct	2439	2458	SEQ ID NO:	4895	aagattaatccgcctgct	8570	8589	3
SEQ ID NO:	3567	ctggagagagagctgggt	2455	2474	SEQ ID NO:	4896	accactggagactaccag	12527	12546	3
SEQ ID NO:	3568	ggagctggttggtccagt	2464	2483	SEQ ID NO:	4897	actggtggcaaaacctcc	2734	2753	3
SEQ ID NO:	3569	gtgttggtccagctcca	2469	2488	SEQ ID NO:	4898	ggagaagccacactcaa	10771	10790	3
SEQ ID NO:	3570	cagctccagctccctcc	2479	2498	SEQ ID NO:	4899	ctggtccctcccaactg	3538	3557	3
SEQ ID NO:	3571	ctccagctccctccctcc	2483	2502	SEQ ID NO:	4900	ggagctggtccctccag	2672	2691	3
SEQ ID NO:	3572	ctgggaagaagctctgca	2501	2520	SEQ ID NO:	4901	tcagaagctacactccag	7939	7958	3
SEQ ID NO:	3573	gagctcagcaggaaggtc	2561	2580	SEQ ID NO:	4902	agccaggaagtgagatctc	3514	3533	3
SEQ ID NO:	3574	agaagagctttttcttc	2582	2601	SEQ ID NO:	4903	gaaggtctcgaggagctt	3835	3854	3
SEQ ID NO:	3575	ctttttcttcaactac	2590	2609	SEQ ID NO:	4904	gagcttcaacactaaag	6107	6126	3
SEQ ID NO:	3576	catctctcatggagagctc	2605	2624	SEQ ID NO:	4905	ggacacttccaaatgtatg	10718	10737	3
SEQ ID NO:	3577	cttcagtgagagctcttt	2608	2627	SEQ ID NO:	4906	aagattaatctggagag	12281	12300	3
SEQ ID NO:	3578	aagctcttggaactcccca	2618	2637	SEQ ID NO:	4907	gggctcggtcagcpcatt	5737	5756	3
SEQ ID NO:	3579	gctcttggaactccacctg	2621	2640	SEQ ID NO:	4908	cagctcagcagctcagcgc	5383	5402	3
SEQ ID NO:	3580	caaggctggagtaaaactg	2692	2711	SEQ ID NO:	4909	cagctcagcagcagcagctg	5080	5099	3
SEQ ID NO:	3581	ggagtaaaactggagaatla	2698	2717	SEQ ID NO:	4910	taactccaagcagcagctcca	3050	3078	3
SEQ ID NO:	3582	ggagtagccaacagcgag	2710	2729	SEQ ID NO:	4911	ctgcatctcagcagcttcc	4106	4125	3
SEQ ID NO:	3583	ttgtgacaatatgggca	2765	2784	SEQ ID NO:	4912	gtcctgtgtacacacaaa	11236	11255	3
SEQ ID NO:	3584	gtgacaatatgggcatc	2767	2786	SEQ ID NO:	4913	gtatggtctctacgcgaca	4385	4404	3
SEQ ID NO:	3585	ggactctcgtcagagggg	2794	2813	SEQ ID NO:	4914	ccaagggccaagggtgtcc	12341	12360	3
SEQ ID NO:	3586	gtgggttcagagtgacac	2808	2827	SEQ ID NO:	4915	gtgtctgacactctccac	4179	4198	3
SEQ ID NO:	3587	ttccagagctcgggtctg	2834	2853	SEQ ID NO:	4916	ccaggaatctgtaccagaa	12562	12581	3
SEQ ID NO:	3588	agctgggtcgtcagagctca	2841	2860	SEQ ID NO:	4917	ggagaaactcagctgact	4807	4826	3
SEQ ID NO:	3589	tcgggtcgtcgggctcag	2843	2862	SEQ ID NO:	4918	ctcaggaagccaattccga	7639	7658	3
SEQ ID NO:	3590	aaagctcgggagctcag	2869	2888	SEQ ID NO:	4919	ctctccagacactgatttt	7951	7970	3
SEQ ID NO:	3591	aaagctcgaatttattcat	2879	2898	SEQ ID NO:	4920	gaaatttcaaatgtgctt	6269	6288	3
SEQ ID NO:	3592	gagaccagctcagctctc	2908	2927	SEQ ID NO:	4921	gagctctcaggaagctctc	13214	13233	3
SEQ ID NO:	3593	gcaacacattacatttgt	2934	2953	SEQ ID NO:	4922	accagctcagatattgttc	10191	10210	3
SEQ ID NO:	3594	acattacatttgtctcta	2938	2957	SEQ ID NO:	4923	agaaatgtgactaaatgt	11889	11908	3
SEQ ID NO:	3595	cattacatttgtctctac	2940	2959	SEQ ID NO:	4924	gtagctcaggaatcaatg	7106	7125	3
SEQ ID NO:	3596	aaacggggagtgatccacc	2964	2983	SEQ ID NO:	4925	ggggagtaacctgaagttt	3205	3224	3
SEQ ID NO:	3597	attgagaacaggcagctcc	2987	3006	SEQ ID NO:	4926	aggaaagcgcacactcaat	12031	12050	3
SEQ ID NO:	3598	tgagaaacaggcagctctg	2989	3008	SEQ ID NO:	4927	ccagctctcccaactctca	8341	8360	3
SEQ ID NO:	3599	ctgacactcagggtctac	3043	3062	SEQ ID NO:	4928	cttaagaaatacagagacg	6440	6459	3
SEQ ID NO:	3600	ctccacagactcggctctc	3074	3093	SEQ ID NO:	4929	aggacagagactcgttggga	3192	3211	3
SEQ ID NO:	3601	ctgacggggacacacagat	3101	3120	SEQ ID NO:	4930	atctgagtgagaaactcag	12259	12278	3
SEQ ID NO:	3602	tagagctcgaagcagacc	3120	3139	SEQ ID NO:	4931	ggctctctggtgactcta	5144	5163	3
SEQ ID NO:	3603	ctagagctcagagagag	3175	3194	SEQ ID NO:	4932	ctctcaaaaaagtatag	6549	6568	3

SEQ ID NO: 3604	ctgggtgataccctgaag	3202	3221	SEQ ID NO: 4933	cttcagaagctctcaag	13217	13236	3
SEQ ID NO: 3605	tgtaactcaagcagaag	3222	3241	SEQ ID NO: 4934	ctctacacataatcaca	9530	9549	3
SEQ ID NO: 3606	gaactcaagcagaaggtc	3225	3244	SEQ ID NO: 4935	gcacactagctggaagta	6955	6974	3
SEQ ID NO: 3607	gcagaaggtgcgaagcaga	3233	3252	SEQ ID NO: 4936	gtgtgtggtactcatctg	4091	4110	3
SEQ ID NO: 3608	cagaaggtgcgaagcagac	3234	3253	SEQ ID NO: 4937	gtctgtggtgtccatctg	4090	4109	3
SEQ ID NO: 3609	gtatgacctgtccagta	3288	3307	SEQ ID NO: 4938	tcaccaacggagaaacatac	10851	10870	3
SEQ ID NO: 3610	taagacctgtccagttaa	3289	3308	SEQ ID NO: 4939	ttcacaacggagaaacata	10850	10869	3
SEQ ID NO: 3611	gaagttcaaaattccggait	3305	3324	SEQ ID NO: 4940	aatctcaagctttctctc	10052	10071	3
SEQ ID NO: 3612	gagggcacaacgtcttaca	3371	3390	SEQ ID NO: 4941	gttaacactgttcgcctc	4215	4234	3
SEQ ID NO: 3613	ggggcacaacgtcttaca	3372	3391	SEQ ID NO: 4942	ctgttaggacacacacct	4062	4081	3
SEQ ID NO: 3614	gaactcaccctgacattca	3390	3409	SEQ ID NO: 4943	gaaaattcaatcacaaagtc	9076	9095	3
SEQ ID NO: 3615	ctggacatcagaaacaaga	3398	3417	SEQ ID NO: 4944	ctctttcttcagccagc	9226	9245	3
SEQ ID NO: 3616	ctatggcgacctaagttg	3435	3454	SEQ ID NO: 4945	caactcgacacatatatga	6635	6654	3
SEQ ID NO: 3617	tgggcgacctaagtttga	3438	3457	SEQ ID NO: 4946	ctactcattatacctccca	6316	6335	3
SEQ ID NO: 3618	agttgtgacacaagaagga	3449	3468	SEQ ID NO: 4947	ctcttttccaattgaact	13838	13857	3
SEQ ID NO: 3619	gacacacaagaagaagaaga	3454	3473	SEQ ID NO: 4948	ctctctctctcatctgtca	10220	10239	3
SEQ ID NO: 3620	gacacacaagaagaagaaga	3455	3474	SEQ ID NO: 4949	ctctctctctcatctgtc	10219	10238	3
SEQ ID NO: 3621	gagaagaagaagaatcaag	3463	3482	SEQ ID NO: 4950	ctgtctgacctcagcttcc	11348	11367	3
SEQ ID NO: 3622	aaatcaaggggtgtattt	3473	3492	SEQ ID NO: 4951	aaatcttatggggatttt	7084	7103	3
SEQ ID NO: 3623	ccataccocgtttccaag	3491	3510	SEQ ID NO: 4952	cttggtatcaaaattgtga	6858	6877	3
SEQ ID NO: 3624	tgcaagcagagccagaag	3504	3523	SEQ ID NO: 4953	cttcagggaacacatactca	5185	5204	3
SEQ ID NO: 3625	cagaagccagaagtgagat	3510	3529	SEQ ID NO: 4954	atctatgtccactctctg	5633	5652	3
SEQ ID NO: 3626	tgagatctctgcccactgg	3523	3542	SEQ ID NO: 4955	ccagcttcccacatctca	8341	8360	3
SEQ ID NO: 3627	gtgtgcctgccaactcgt	3540	3559	SEQ ID NO: 4956	agcacatgtgaactggacc	13947	13966	3
SEQ ID NO: 3628	gtcttccaaatggacatc	3555	3574	SEQ ID NO: 4957	gagtttatcagtcagagac	9701	9720	3
SEQ ID NO: 3629	ggagctatctgdcacgc	3567	3586	SEQ ID NO: 4958	gctgcagtgcccgcttcca	8167	8186	3
SEQ ID NO: 3630	gtctacagcttatggctcca	3578	3597	SEQ ID NO: 4959	tgaggagcatctcttagc	8211	8230	3
SEQ ID NO: 3631	gggtgcatggcattatgat	3610	3629	SEQ ID NO: 4960	atcacaatagttttcaac	8947	8966	3
SEQ ID NO: 3632	ggagaagatggaattgaa	3631	3650	SEQ ID NO: 4961	ttcaacgtatccgtctct	7713	7732	3
SEQ ID NO: 3633	caggcaccagtgtagatc	3657	3676	SEQ ID NO: 4962	gtatgtctaaagactctg	3736	3755	3
SEQ ID NO: 3634	gaactccaatttccctgtg	3685	3704	SEQ ID NO: 4963	cacaatgcaaaattcagtc	5195	5214	3
SEQ ID NO: 3635	gtccctcaaacacacatga	3764	3783	SEQ ID NO: 4964	ctacaaggaggttagggac	12777	12796	3
SEQ ID NO: 3636	caaacagacatgacttcc	3770	3789	SEQ ID NO: 4965	ggaactcaatttcaattg	7022	7041	3
SEQ ID NO: 3637	atagttgcaatgacatc	3809	3828	SEQ ID NO: 4966	atgatttgaataatgcat	6693	6712	3
SEQ ID NO: 3638	gttcagagagcctctggg	3829	3848	SEQ ID NO: 4967	cccaagaggtatttaaacg	12957	12976	3
SEQ ID NO: 3639	ggagttcaaccctccagaac	3895	3914	SEQ ID NO: 4968	gttctactcatttaaccctc	6314	6333	3
SEQ ID NO: 3640	ggaacaactcttctataaa	3940	3959	SEQ ID NO: 4969	tttctaaatggaacttct	12173	12192	3
SEQ ID NO: 3641	aaacctcttctataaaag	3942	3961	SEQ ID NO: 4970	cttgaataaattcttttt	9213	9232	3
SEQ ID NO: 3642	aaaagcgatggccgggtc	3955	3974	SEQ ID NO: 4971	gaccttgcagaatatttt	6343	6362	3
SEQ ID NO: 3643	gtcaaatataccttgaaca	3971	3990	SEQ ID NO: 4972	gttaacaattcttgac	7363	7382	3
SEQ ID NO: 3644	tgacaacagaacagcttga	3984	4003	SEQ ID NO: 4973	ttcaagttctgactctca	8310	8329	3
SEQ ID NO: 3645	gtttgaaatttgagattc	3995	4014	SEQ ID NO: 4974	ggaatctgtctctccaact	9047	9066	3
SEQ ID NO: 3646	gttgaattgagatttcc	3996	4015	SEQ ID NO: 4975	gggaataacaaagctcaaac	10454	10473	3
SEQ ID NO: 3647	tgaaattgagatttctt	3998	4017	SEQ ID NO: 4976	agggaaaagcgccactcaa	12030	12049	3
SEQ ID NO: 3648	ctaaagatgttagagactg	4046	4065	SEQ ID NO: 4977	cagttgaccacaaagcttag	10545	10564	3
SEQ ID NO: 3649	atgttagagactgttagga	4052	4071	SEQ ID NO: 4978	ctctaacaacctctgaact	8073	8092	3
SEQ ID NO: 3650	cagcctcactctcaagtc	4074	4093	SEQ ID NO: 4979	ggaactcttagtcaggctg	8813	8832	3

SEQ ID NO: 3651	agccctccacttcaagtct	4075	4094	SEQ ID NO: 4980	agacatgctgggtggct	5728	5747	3
SEQ ID NO: 3652	ccatctgcacatctcgagag	4102	4121	SEQ ID NO: 4981	ctctcaaatgacatgtag	5330	5349	3
SEQ ID NO: 3653	attcccaagtgtgtataaac	4142	4161	SEQ ID NO: 4982	gttgtagaacgcccaagaat	6254	6273	3
SEQ ID NO: 3654	taactgcaagtgcctctc	4156	4175	SEQ ID NO: 4983	gagatcaagagcaactgtga	8843	8862	3
SEQ ID NO: 3655	gggtgtctagacctctca	4178	4197	SEQ ID NO: 4984	tggaaacctctccctacc	4735	4754	3
SEQ ID NO: 3656	ctccagatgtctctacagc	4192	4211	SEQ ID NO: 4985	cgctgtgaactcaaaagag	5588	5607	3
SEQ ID NO: 3657	gcgaagtgtctacagcaac	4195	4214	SEQ ID NO: 4986	gtgtgccaccatcatctgtg	11671	11690	3
SEQ ID NO: 3658	acgaagtgtctacagcaact	4198	4215	SEQ ID NO: 4987	agttgccaccatcatctgt	11670	11689	3
SEQ ID NO: 3659	ttctacagtggtgtgcaaca	4232	4251	SEQ ID NO: 4988	gttagttgtctcttaagga	13359	13378	3
SEQ ID NO: 3660	gtgtccacatgaaagctgt	4280	4299	SEQ ID NO: 4989	cagcaagtcactgcagaagc	8611	8630	3
SEQ ID NO: 3661	gaaggctgactctgtgggt	4291	4310	SEQ ID NO: 4990	aacctatgccttaattctc	13189	13188	3
SEQ ID NO: 3662	gtgtgtgacctgtcttcc	4303	4322	SEQ ID NO: 4991	ggaagtgtaaacaacaca	6985	6984	3
SEQ ID NO: 3663	ctgtcttctctacaagtgtg	4312	4331	SEQ ID NO: 4992	cacacctgacattgcagg	11088	11107	3
SEQ ID NO: 3664	ctgtcttctctacaagtgtc	4313	4332	SEQ ID NO: 4993	gcacacctgacattgcag	11087	11106	3
SEQ ID NO: 3665	ttctacaagtgtcaaggat	4319	4338	SEQ ID NO: 4994	atccctgggtctgaagga	8577	8596	3
SEQ ID NO: 3666	atgacaccaagaataactgt	4352	4371	SEQ ID NO: 4995	agcttcgtgtgtgcctcata	9984	10003	3
SEQ ID NO: 3667	atgacaccaagaataactgtc	4353	4372	SEQ ID NO: 4996	gagcttcgtgtgtgcctcat	9983	10002	3
SEQ ID NO: 3668	gaatagctctacactatca	4363	4382	SEQ ID NO: 4997	gtattatctgaattcttc	6487	6506	3
SEQ ID NO: 3669	tttagtattcgaatatca	4406	4425	SEQ ID NO: 4998	gatttcatgatttgaaa	8685	8704	3
SEQ ID NO: 3670	gatttgaatatcaaatca	4412	4431	SEQ ID NO: 4999	gaagtagctgagaaaatc	7102	7121	3
SEQ ID NO: 3671	gaacacaaccagctctcaaa	4449	4468	SEQ ID NO: 5000	ttgaaaaattctcttctc	9214	9233	3
SEQ ID NO: 3672	cccagctctcaaaagggtta	4456	4475	SEQ ID NO: 5001	taattcaattactctctgg	11302	11321	3
SEQ ID NO: 3673	ctcaaaaggtttactaata	4462	4481	SEQ ID NO: 5002	tatcaaaactgagttgag	12231	12250	3
SEQ ID NO: 3674	ttcaaaaggtttactaataat	4463	4482	SEQ ID NO: 5003	atattcaaaactgagttga	12230	12249	3
SEQ ID NO: 3675	aaaaggtttactaataattc	4465	4484	SEQ ID NO: 5004	gaatttgaaggttgtttt	9280	9299	3
SEQ ID NO: 3676	gaacacagcatttgtgtc	4543	4562	SEQ ID NO: 5005	gcagacatctctgtgttc	11214	11233	3
SEQ ID NO: 3677	atttgttgtcaaaagaagt	4551	4570	SEQ ID NO: 5006	acttaaaaaataataaat	8022	8041	3
SEQ ID NO: 3678	ttcaagattgtggcgagtt	4569	4588	SEQ ID NO: 5007	aacttcaagtaactgtga	13422	13441	3
SEQ ID NO: 3679	ttcagagctctcttgttct	4586	4605	SEQ ID NO: 5008	gaagatgtgcaaatgtga	11995	12014	3
SEQ ID NO: 3680	cagagctctcttgttctat	4588	4607	SEQ ID NO: 5009	atagcatggaactctctg	8873	8892	3
SEQ ID NO: 3681	gtgtcaaaagtcacatatgg	4605	4624	SEQ ID NO: 5010	ccatttgtgagatcaggcat	9245	9264	3
SEQ ID NO: 3682	gcacatatgtgcctcttgtc	4614	4633	SEQ ID NO: 5011	caagttgtgcaagtaagtcg	9372	9391	3
SEQ ID NO: 3683	gagttcaacctgaggttta	4667	4686	SEQ ID NO: 5012	taaaagtgcaacttactac	6190	6209	3
SEQ ID NO: 3684	agttcaacctgaggttttaa	4668	4687	SEQ ID NO: 5013	taacagggaagatagact	9308	9327	3
SEQ ID NO: 3685	ctcactctcgaagaccacaa	4692	4711	SEQ ID NO: 5014	ttgtgcaagtaagtcagg	9376	9395	3
SEQ ID NO: 3686	gaagatgtgaacctctctcc	4730	4749	SEQ ID NO: 5015	gggaagaagaaggacattc	12291	12310	3
SEQ ID NO: 3687	gtatctgcaaggtgtgacct	4762	4781	SEQ ID NO: 5016	gatgagggaactcagatca	12283	12281	3
SEQ ID NO: 3688	gatctgtcaaaagtgcacata	4763	4782	SEQ ID NO: 5017	gtatgagggaactcagatc	12282	12281	3
SEQ ID NO: 3689	gtcttccctaaagttagaga	4793	4812	SEQ ID NO: 5018	tttgtgtgtctggaanaagc	5977	5996	3
SEQ ID NO: 3690	gtatgagaactacgagctg	4804	4823	SEQ ID NO: 5019	cagctttaagagacacatac	8920	8939	3
SEQ ID NO: 3691	ttctacaagaatggaatgat	4868	4887	SEQ ID NO: 5020	ttattttccaactaataga	13032	13051	3
SEQ ID NO: 3692	gtctgtgtgttgtgaatact	4907	4926	SEQ ID NO: 5021	gtatcacagaanaactgcag	6901	6920	3
SEQ ID NO: 3693	ttcattgaggtctctcagcc	4940	4959	SEQ ID NO: 5022	ggctctatctgtgaatga	5348	5367	3
SEQ ID NO: 3694	ttctgtgatactaaatttc	4963	4982	SEQ ID NO: 5023	gggaaggacaaggccagaa	12549	12568	3
SEQ ID NO: 3695	tcattgtcttgtagttaaat	4981	5000	SEQ ID NO: 5024	atttttttctctgcattg	10103	10122	3
SEQ ID NO: 3696	ctttaggcactgacaaat	5007	5026	SEQ ID NO: 5025	atttttttcaaggttaasga	14019	14038	3
SEQ ID NO: 3697	tcaaggcgacactaaggat	5040	5059	SEQ ID NO: 5026	atccatgatctacatttgt	6794	6813	3

SEQ ID NO: 3689	gcaacgaccaactgaag	5083	5102	SEQ ID NO: 5027	cttcagggaacacaatgca	5185	5204	3
SEQ ID NO: 3689	caactgaagtgatgctc	5092	5111	SEQ ID NO: 5028	gagatgagagatgcgttg	6239	6258	3
SEQ ID NO: 3700	ctgtgagatgagctgaat	5116	5135	SEQ ID NO: 5029	attctctttcttttcagc	9222	9241	3
SEQ ID NO: 3701	gcagagctggcctctctg	5138	5154	SEQ ID NO: 5030	cagatatacagaaaaactgc	6899	6918	3
SEQ ID NO: 3702	ctctggggcactatgaa	5148	5167	SEQ ID NO: 5031	ttactcaattgggagaga	6499	6518	3
SEQ ID NO: 3703	ctctggggcactatgaaat	5150	5168	SEQ ID NO: 5032	attgttaagaanaatacaga	6436	6455	3
SEQ ID NO: 3704	acacacatgcaaaactcag	5193	5212	SEQ ID NO: 5033	ctgaagcatlaaaaactgtt	7506	7525	3
SEQ ID NO: 3705	ctcacagagctatcactgg	5231	5250	SEQ ID NO: 5034	ccagatgctgaacagctgag	8149	8168	3
SEQ ID NO: 3706	gggaagtgcttatcaggc	5247	5266	SEQ ID NO: 5035	gcctactgttccatgtacca	11356	11375	3
SEQ ID NO: 3707	ctcaagctcagcagaaga	5303	5322	SEQ ID NO: 5036	ctctcagtgagaaatgaa	11977	11996	3
SEQ ID NO: 3708	atgacatgaggggctcat	5336	5355	SEQ ID NO: 5037	atgattatctgaattcatt	6486	6505	3
SEQ ID NO: 3709	gctcatatgctgaatgaa	5348	5368	SEQ ID NO: 5038	ttcagcatgacatgagc	5746	5765	3
SEQ ID NO: 3710	atagctgaatgaatt	5353	5372	SEQ ID NO: 5039	aaatagctatgtcctaatt	6702	6721	3
SEQ ID NO: 3711	ctgacatgtgcaggctta	5386	5405	SEQ ID NO: 5040	taagaaccagagatcaga	10996	11015	3
SEQ ID NO: 3712	gaacattgcaggcttatca	5389	5408	SEQ ID NO: 5041	tgatatgcagctgaggctt	12480	12509	3
SEQ ID NO: 3713	gcagcgtttacatcgagc	5395	5414	SEQ ID NO: 5042	gtctctgattccacatgca	11852	11871	3
SEQ ID NO: 3714	caaaaactgcacaacattt	5420	5439	SEQ ID NO: 5043	aaaattctgtcacatgtga	7370	7389	3
SEQ ID NO: 3715	tttaccagctgcagcaagt	5435	5454	SEQ ID NO: 5044	acttaaaaataataaaaat	8022	8041	3
SEQ ID NO: 3716	ctctgacaagttttataag	5443	5462	SEQ ID NO: 5045	cttacttgtaattccaagag	10674	10693	3
SEQ ID NO: 3717	tttaattacacgtcacagc	5468	5487	SEQ ID NO: 5046	gctgcagctgctgtgtaac	5578	5597	3
SEQ ID NO: 3718	ttctctgtglaactcttta	5491	5510	SEQ ID NO: 5047	taaaagattacttggagaa	7275	7294	3
SEQ ID NO: 3719	ctcaaaaggagcctaccaa	5596	5615	SEQ ID NO: 5048	ttggcagatgaagtgtcagg	9376	9395	3
SEQ ID NO: 3720	aaaaggagcctacacaaat	5599	5618	SEQ ID NO: 5049	atttacaattgttctttt	6271	6290	3
SEQ ID NO: 3721	ggagcctacacaaaataat	5602	5621	SEQ ID NO: 5050	attactcgtatgttctct	10127	10146	3
SEQ ID NO: 3722	ataatgaataaaacacat	5616	5635	SEQ ID NO: 5051	atgloaaacactttgttat	7065	7084	3
SEQ ID NO: 3723	aaaaacacatctatgcac	5626	5645	SEQ ID NO: 5052	galgagatgacgcacttt	12158	12177	3
SEQ ID NO: 3724	gtctaaggttcagggtgtg	5686	5705	SEQ ID NO: 5053	cacaagtgatttccacagca	9087	9106	3
SEQ ID NO: 3725	gagtttagcctcaggtcca	5705	5724	SEQ ID NO: 5054	tgaggctgactcagagactc	7450	7469	3
SEQ ID NO: 3726	gtggctctcagccatgac	5740	5759	SEQ ID NO: 5055	gtcagtggaagtctccagc	8596	8615	3
SEQ ID NO: 3727	atttcagcaaatgctcttcg	5790	5809	SEQ ID NO: 5056	cggagcatgggagtgaaat	8628	8647	3
SEQ ID NO: 3728	tttcagcaaatgtcttcgt	5791	5810	SEQ ID NO: 5057	acggcagctgggagtgaaa	8627	8646	3
SEQ ID NO: 3729	tttcagcaaatgtcttcgt	5792	5811	SEQ ID NO: 5058	aaacggcagctgggagtgaa	8626	8645	3
SEQ ID NO: 3730	cagcaaatgtcttcgttct	5794	5813	SEQ ID NO: 5059	agaagtgcttcaaaagctg	12412	12431	3
SEQ ID NO: 3731	tgctctcgttctgttaag	5800	5819	SEQ ID NO: 5060	catctcaattgggagagaca	6501	6520	3
SEQ ID NO: 3732	gtcttcgttctgttaag	5801	5820	SEQ ID NO: 5061	ccattcagctctccaagac	12975	12994	3
SEQ ID NO: 3733	atgggaaactcgtctctctg	5859	5878	SEQ ID NO: 5062	cagataaaaactcaacat	12213	12232	3
SEQ ID NO: 3734	ggagaaacatactgggcagc	5879	5898	SEQ ID NO: 5063	gctgttttgaagactcttc	1088	1107	3
SEQ ID NO: 3735	gttgaacgcagaacactctg	5814	5833	SEQ ID NO: 5064	cagaattcataatccaac	8274	8293	3
SEQ ID NO: 3736	gtctagtgaaaagcatcagt	5883	6002	SEQ ID NO: 5065	actgcagcagattttcagac	13612	13631	3
SEQ ID NO: 3737	agcatcagtcagctcttg	5993	6012	SEQ ID NO: 5066	ccaagaacctgttagttgt	13351	13370	3
SEQ ID NO: 3738	tgaaacacaaagtcagtc	6009	6028	SEQ ID NO: 5067	gcacataaatatgatcaa	6418	6437	3
SEQ ID NO: 3739	gcagcagcagcactgtgaaa	6046	6065	SEQ ID NO: 5068	tttcagatggcatgctdgc	11610	11629	3
SEQ ID NO: 3740	gaacacaaagcacaattt	6061	6080	SEQ ID NO: 5069	aaatcccatacagggttct	8037	8056	3
SEQ ID NO: 3741	acaatgaatcacagcagga	6084	6103	SEQ ID NO: 5070	tccttggctgtgtctgtt	9682	9701	3
SEQ ID NO: 3742	cttgagtgctatacaactc	6103	6122	SEQ ID NO: 5071	agtgaagttctccagcaga	8599	8618	3
SEQ ID NO: 3743	tggtgtggagcttacttg	6132	6151	SEQ ID NO: 5072	ccaagaattcataataccaa	8273	8292	3
SEQ ID NO: 3744	caatttactcagtgagcc	6198	6217	SEQ ID NO: 5073	ggctattgattagagtg	6988	7007	3

SEQ ID NO: 3745	tttagagatgagagatgcc	6235	6254	SEQ ID NO: 5074	ggcatgatgctcattaaa	9177	9196	3
SEQ ID NO: 3746	gagaagccccaagaattta	6257	6276	SEQ ID NO: 5075	taaaagccattcagctctc	12970	12989	3
SEQ ID NO: 3747	caattgtgtctgttataaa	6276	6295	SEQ ID NO: 5076	ttaaacaggatcagatattg	10187	10206	3
SEQ ID NO: 3748	ttttgaaagatgatataa	6286	6305	SEQ ID NO: 5077	ttatgtcgtatccaaaaa	13655	13674	3
SEQ ID NO: 3749	tgttaaglatgatataaaa	6288	6307	SEQ ID NO: 5078	ttttagagggaatcgacaaa	6358	6377	3
SEQ ID NO: 3750	ttcaactcattaaacctcc	6315	6334	SEQ ID NO: 5079	ggggaaaaaacagcgtgaa	9576	9595	3
SEQ ID NO: 3751	tttgagacctgcagagaa	6337	6356	SEQ ID NO: 5080	ttctctctatlgggaaaaa	9566	9585	3
SEQ ID NO: 3752	acctgtcaagaatattttg	6344	6363	SEQ ID NO: 5081	caaaagaagcccaagaggt	12948	12967	3
SEQ ID NO: 3753	tcaatattgatacaatttgg	6423	6442	SEQ ID NO: 5082	caaaagcagatattgttga	11829	11848	3
SEQ ID NO: 3754	caagacagcgtctgggaaaa	6451	6470	SEQ ID NO: 5083	ttttcagaccaactctctg	13622	13641	3
SEQ ID NO: 3755	ctgtggaaaactccacag	6460	6479	SEQ ID NO: 5084	ctgtctctgtcagccagg	7724	7743	3
SEQ ID NO: 3756	actccacagcagaagctaat	6469	6488	SEQ ID NO: 5085	ttaactactctcttgcagtt	12869	12888	3
SEQ ID NO: 3757	taattcattcaattgggaga	6497	6516	SEQ ID NO: 5086	ctctctctctcatgtgaatt	10479	10498	3
SEQ ID NO: 3758	tcaattgggagagacagaag	6503	6522	SEQ ID NO: 5087	ctgttgagtgccagtttga	11808	11827	3
SEQ ID NO: 3759	aggagaacatgactgctct	6534	6553	SEQ ID NO: 5088	gagacttatgggatttctc	11163	11182	3
SEQ ID NO: 3760	actgactgctctcaacaaa	6541	6560	SEQ ID NO: 5089	tttggcagctatatacagt	8380	8399	3
SEQ ID NO: 3761	gactgctctcaacaaaaga	6544	6563	SEQ ID NO: 5090	cttttgaggtttacagtc	9695	9714	3
SEQ ID NO: 3762	cagacatatgatacaaat	6641	6660	SEQ ID NO: 5091	tattgatatcagaagctctg	1933	1952	3
SEQ ID NO: 3763	aattgtatcatatattaaa	6657	6676	SEQ ID NO: 5092	taaaagaaatctcaatt	13815	13834	3
SEQ ID NO: 3764	tatgatttacaatgattga	6683	6702	SEQ ID NO: 5093	taagtattatataccata	13128	13147	3
SEQ ID NO: 3765	tttgaataatgactatgct	6697	6716	SEQ ID NO: 5094	gacacagacaaaattcaaa	13864	13883	3
SEQ ID NO: 3766	ttgaaaatagctattgcta	6698	6717	SEQ ID NO: 5095	gacacagacaaaattcaaa	13863	13882	3
SEQ ID NO: 3767	taatagtattgtctaattat	6703	6722	SEQ ID NO: 5096	ataataatggagctttatt	14084	14103	3
SEQ ID NO: 3768	tattattgtatgaatcattg	6719	6738	SEQ ID NO: 5097	taatacagaaatctataat	8268	8287	3
SEQ ID NO: 3769	aaagtcctgatgagcacta	6747	6766	SEQ ID NO: 5098	tggtgttaactctctctt	12864	12883	3
SEQ ID NO: 3770	aagctgtatgagcactat	6748	6767	SEQ ID NO: 5099	atagcaacactaaataact	8769	8788	3
SEQ ID NO: 3771	ttagtgagcactatcata	6753	6772	SEQ ID NO: 5100	atataccaagatgagatcaa	13101	13120	3
SEQ ID NO: 3772	taattttgtataaacaat	6777	6796	SEQ ID NO: 5101	tattgagatctccatcata	11702	11721	3
SEQ ID NO: 3773	ttttgataaaaacaatcca	6780	6799	SEQ ID NO: 5102	ggagtgccaggttgaaaa	11810	11829	3
SEQ ID NO: 3774	acattgtttattgaaatt	6805	6824	SEQ ID NO: 5103	attctctaaagctgtgagt	11175	11194	3
SEQ ID NO: 3775	tattgatttaacaagaattg	6824	6843	SEQ ID NO: 5104	caactgtctccagttgtcaat	9871	9890	3
SEQ ID NO: 3776	attttaacaaaagtggaag	6828	6847	SEQ ID NO: 5105	ctctaaagacttaaaaaat	8014	8033	3
SEQ ID NO: 3777	taatacagaatccagatata	6888	6907	SEQ ID NO: 5106	gtaccataagccataatt	10088	10107	3
SEQ ID NO: 3778	gaatccagatatacaagaaa	6894	6913	SEQ ID NO: 5107	tttctaaacttgaaatct	9065	9084	3
SEQ ID NO: 3779	tttagagacatacatcagaa	6924	6943	SEQ ID NO: 5108	ttcttaaacactctcttaa	9491	9510	3
SEQ ID NO: 3780	atccagcaactcagctggaa	6950	6969	SEQ ID NO: 5109	ttccaacttccctgttgat	3688	3707	3
SEQ ID NO: 3781	tgagcatgtcaaacacttt	7060	7079	SEQ ID NO: 5110	aaagtgcacttttactca	6191	6210	3
SEQ ID NO: 3782	agcatgtcaaacactttg	7061	7080	SEQ ID NO: 5111	caaatgacatgatgggtct	5334	5353	3
SEQ ID NO: 3783	aaacactttgtataaatc	7070	7089	SEQ ID NO: 5112	gtattatataccattatttt	13133	13152	3
SEQ ID NO: 3784	tgagaaaatcaatgccttc	7111	7130	SEQ ID NO: 5113	gaaggaaaagcgacactca	12029	12048	3
SEQ ID NO: 3785	tatgaagtagaccacaacaa	7160	7179	SEQ ID NO: 5114	ttgtggagggtatgcata	10331	10350	3
SEQ ID NO: 3786	taagttagaccacaatacca	7164	7183	SEQ ID NO: 5115	tgatgaagaatgaagactt	12156	12175	3
SEQ ID NO: 3787	aagttagaaggagacatttg	7223	7242	SEQ ID NO: 5116	gaataccaatctgtgaact	10168	10187	3
SEQ ID NO: 3788	acaagtgaagataaaaagat	7264	7283	SEQ ID NO: 5117	atctaaattcagttcttgt	11334	11353	3
SEQ ID NO: 3789	gaagataaaaagattattg	7271	7290	SEQ ID NO: 5118	caaaatagaaggagaactt	2077	2096	3
SEQ ID NO: 3790	gattactttgagaattttg	7280	7299	SEQ ID NO: 5119	ctaaacttgaaattcaatc	9069	9088	3
SEQ ID NO: 3791	tgagaaattattggatttt	7288	7307	SEQ ID NO: 5120	aaatcgttgaggtgactca	7443	7462	3

SEQ ID NO: 3792	aaattagttgattatg	7292	7311	SEQ ID NO: 5127	caatttgagaatgaatt	10419	10438	3
SEQ ID NO: 3793	tgattattgatgatct	7300	7319	SEQ ID NO: 5128	agatgctagttcttoca	9953	9972	3
SEQ ID NO: 3794	tcattgaagatgttaacaa	7353	7372	SEQ ID NO: 5129	ttgatagtaaaccaatga	7422	7441	3
SEQ ID NO: 3795	catgaagatgttaacaaa	7354	7373	SEQ ID NO: 5124	tttgatagtaaaccaatg	7421	7440	3
SEQ ID NO: 3796	attgaagatgttaacaaat	7355	7374	SEQ ID NO: 5125	attagatgatgttcaat	10495	10514	3
SEQ ID NO: 3797	tgaagatgttaacaaatt	7356	7375	SEQ ID NO: 5126	aaattagatgatgttcaa	10494	10513	3
SEQ ID NO: 3798	gaagatgttaacaaatc	7357	7376	SEQ ID NO: 5127	gaatttagatgttcaat	10493	10512	3
SEQ ID NO: 3799	acattgttgaataagaaatt	7380	7399	SEQ ID NO: 5128	aattccctgaagttgatgt	11487	11506	3
SEQ ID NO: 3800	ttgattaccaccagtttg	7406	7425	SEQ ID NO: 5129	caaatggaacatcccaaaa	8791	8810	3
SEQ ID NO: 3801	caaaatccgtgaggttgact	7441	7460	SEQ ID NO: 5130	agatcccccatacagattg	7972	7991	3
SEQ ID NO: 3802	aaaatccgtgaggtgactc	7442	7461	SEQ ID NO: 5131	gaatgaaatgctgtttttt	8638	8657	3
SEQ ID NO: 3803	agtgactcagagactcaa	7452	7471	SEQ ID NO: 5132	ttgatgatctggaacct	10731	10750	3
SEQ ID NO: 3804	gtgaattcaggctgtgga	7473	7492	SEQ ID NO: 5133	ccaatctctctctttcac	8409	8428	3
SEQ ID NO: 3805	gttcagttgatgtgaaa	7547	7566	SEQ ID NO: 5134	ttcaagcgaatgtcacac	8540	8559	3
SEQ ID NO: 3806	taagttcagcatcttgg	7616	7635	SEQ ID NO: 5135	ccaatgctgaactcttttaa	10173	10192	3
SEQ ID NO: 3807	tgaaggccaattccgaga	7641	7660	SEQ ID NO: 5136	ctctctctctctctca	10213	10232	3
SEQ ID NO: 3808	aatgtatcaaatgacatt	7684	7703	SEQ ID NO: 5137	aagtgaagtcggatcaatt	11021	11040	3
SEQ ID NO: 3809	attcagcaggaactccaac	7700	7719	SEQ ID NO: 5138	gttgagaagcccaagaat	8254	8273	3
SEQ ID NO: 3810	accgtctctgttcagoca	7722	7741	SEQ ID NO: 5139	ggcgaagtaagtcctaggt	9377	9396	3
SEQ ID NO: 3811	ctgtctctgttgatccagc	7723	7742	SEQ ID NO: 5140	ctggactctctagtcagg	8810	8829	3
SEQ ID NO: 3812	gttcagcagggttatagc	7732	7751	SEQ ID NO: 5141	gtcaagaagcagcttgacc	10535	10554	3
SEQ ID NO: 3813	ccaggtttatagcacactt	7738	7757	SEQ ID NO: 5142	aagtcggatctatcttgg	11025	11044	3
SEQ ID NO: 3814	tttatagcacactgtfca	7742	7761	SEQ ID NO: 5143	gacgtctcatcaaaaac	13681	13700	3
SEQ ID NO: 3815	actgttcacctacattct	7753	7772	SEQ ID NO: 5144	gaaaaaaggggattgaagt	10283	10302	3
SEQ ID NO: 3816	ctgattgttggaactctgc	7770	7789	SEQ ID NO: 5145	gcaagttaaagaaaatcag	14026	14045	3
SEQ ID NO: 3817	atgaagcatgttgtaggc	7847	7866	SEQ ID NO: 5146	gcctactctctctctcat	10208	10227	3
SEQ ID NO: 3818	tgaagcattgttgtagca	7848	7867	SEQ ID NO: 5147	tgctcatctctctctca	10207	10226	3
SEQ ID NO: 3819	gggttcactgtctctgaaa	7868	7887	SEQ ID NO: 5148	ttccacctagaaggaacc	8959	8978	3
SEQ ID NO: 3820	ccaagaccatctctgggac	7887	7906	SEQ ID NO: 5149	gtcccccatacagattga	7973	7992	3
SEQ ID NO: 3821	ctctgggaccatgctgcc	7897	7916	SEQ ID NO: 5150	ggcaccagggtctggaagg	13978	13997	3
SEQ ID NO: 3822	ttcaggctcttcgaagaac	7929	7948	SEQ ID NO: 5151	gcttggaaggatctctgaa	9588	9607	3
SEQ ID NO: 3823	tcagataaactccaaga	8004	8023	SEQ ID NO: 5152	ctctatagattccaatgaa	13183	13202	3
SEQ ID NO: 3824	actccaagactaaaaaa	8013	8032	SEQ ID NO: 5153	tttaacaaaagtgtgaagt	6829	6848	3
SEQ ID NO: 3825	atccatcagggttttoca	8039	8058	SEQ ID NO: 5154	tgagaagcaaatctggat	9472	9491	3
SEQ ID NO: 3826	gaattccatctcttaaca	8063	8082	SEQ ID NO: 5155	gttgaagttctctacact	9889	9908	3
SEQ ID NO: 3827	atctctctcttcaaat	8089	8108	SEQ ID NO: 5156	aattccaatttgaagaag	10414	10433	3
SEQ ID NO: 3828	tgaccagatgctgaacag	8145	8164	SEQ ID NO: 5157	ctgttgaagatttatcaa	12932	12951	3
SEQ ID NO: 3829	aataccctgcagactctc	8233	8252	SEQ ID NO: 5158	gaagttctcaatttgatt	8522	8541	3
SEQ ID NO: 3830	tgaacttcacataccagaa	8320	8339	SEQ ID NO: 5159	ttctcttgaaaagggttca	8884	8903	3
SEQ ID NO: 3831	tccagcttcccacatct	8339	8358	SEQ ID NO: 5160	agattctcagatgaaggaa	8921	8940	3
SEQ ID NO: 3832	aagctatacagattctga	8387	8406	SEQ ID NO: 5161	tcagatggcatctgctct	11612	11631	3
SEQ ID NO: 3833	attctgaaatccaacttc	8399	8418	SEQ ID NO: 5162	gagataaacgtgctcgaat	11552	11571	3
SEQ ID NO: 3834	ttccactatagatgcaaat	8422	8441	SEQ ID NO: 5163	atttttgaaaaaacagaaa	9738	9757	3
SEQ ID NO: 3835	caaatgctgacataggaa	8436	8455	SEQ ID NO: 5164	ttccatcaacaactcttgg	9670	9689	3
SEQ ID NO: 3836	agagttccaaattagaagt	8508	8527	SEQ ID NO: 5165	actttacttccaactctc	13410	13429	3
SEQ ID NO: 3837	agaatgtccaaattagaagt	8509	8528	SEQ ID NO: 5166	aactttacttccaactctc	13409	13428	3
SEQ ID NO: 3838	ctcaatttggatttcaa	8527	8546	SEQ ID NO: 5167	ttgatctcttttttgaaga	11537	11556	3

SEQ ID NO: 3839	caattttgattttcaagca	8530	8549	SEQ ID NO: 5168	gctgaatccaaaagaitg	13660	13679	3
SEQ ID NO: 3840	aatgcacaaactccaaccc	8549	8568	SEQ ID NO: 5169	gggttatcaaggggcatt	12460	12479	3
SEQ ID NO: 3841	agttctccagcaagatcact	8604	8623	SEQ ID NO: 5170	aggttccactgtgtcaaat	11388	11407	3
SEQ ID NO: 3842	agtccttgagaaacggagca	8616	8635	SEQ ID NO: 5171	gctccaggaggaactct	13780	13799	3
SEQ ID NO: 3843	caaacacagtgggcaagt	8678	8697	SEQ ID NO: 5172	aactctcaagtcagatgta	13422	13441	3
SEQ ID NO: 3844	acatacatgtcttaccctgga	8751	8770	SEQ ID NO: 5173	ttccatctgaataatgt	13380	13399	3
SEQ ID NO: 3845	gtgatagcaacacataaat	8765	8784	SEQ ID NO: 5174	attcttgaaactcccccag	12702	12721	3
SEQ ID NO: 3846	ctgacctgagcaacagagat	8829	8848	SEQ ID NO: 5175	atctgtatgaggaaactcag	12259	12278	3
SEQ ID NO: 3847	agatgagggaacacatgaa	8929	8948	SEQ ID NO: 5176	ttcatgttccatgaatact	10038	10057	3
SEQ ID NO: 3848	caacatttttcaaaccttga	9060	9079	SEQ ID NO: 5177	tcaggatataactgttttga	12618	12637	3
SEQ ID NO: 3849	ttctaacttgaataatcaa	9067	9086	SEQ ID NO: 5178	ttgatgatgctgtcagaag	7308	7327	3
SEQ ID NO: 3850	gaattcaatcacaacagtg	9077	9096	SEQ ID NO: 5179	cgacgaagaaaataatttc	13566	13585	3
SEQ ID NO: 3851	cadgttttgagagaaggaa	9141	9160	SEQ ID NO: 5180	ttccagaagaagcagccagt	12506	12525	3
SEQ ID NO: 3852	actgttttgagagaaggaa	9142	9161	SEQ ID NO: 5181	cttccccaaaagagaccagt	2898	2917	3
SEQ ID NO: 3853	aattctcttttctttcag	9221	9240	SEQ ID NO: 5182	gttactacttgaaaaatt	13638	13657	3
SEQ ID NO: 3854	ttcttttcagccagccat	9230	9249	SEQ ID NO: 5183	atgaaaaggaaggaagagaa	13484	13513	3
SEQ ID NO: 3855	ttgaaagttctgttttcca	9283	9302	SEQ ID NO: 5184	tggaagttctagtgcaaaa	10380	10399	3
SEQ ID NO: 3856	gaagggaagatgactctct	9312	9331	SEQ ID NO: 5185	aggaccctttcaaatcttcg	9848	9867	3
SEQ ID NO: 3857	ataagtcacaaccaaatatt	9405	9424	SEQ ID NO: 5186	aaatcaggatctgtagtatt	14038	14057	3
SEQ ID NO: 3858	acacagcagaacattatgaa	9435	9454	SEQ ID NO: 5187	ttcatctgaataatgt	13380	13399	3
SEQ ID NO: 3859	gggaataatctgagaagca	9463	9482	SEQ ID NO: 5188	gtctggaattgttcaatct	11734	11753	3
SEQ ID NO: 3860	agcaaatctggaattctta	9478	9497	SEQ ID NO: 5189	taagttctctgtactctgt	11719	11738	3
SEQ ID NO: 3861	ttcttaacaactctctgaa	9502	9521	SEQ ID NO: 5190	ttcaaaaagagctctcagga	13206	13225	3
SEQ ID NO: 3862	ttacaactttctgaaattg	9505	9524	SEQ ID NO: 5191	catgttattgaagtgtaaa	9621	9640	3
SEQ ID NO: 3863	acacataatcacaaactcc	9534	9553	SEQ ID NO: 5192	ggagacagcatcttctgt	11211	11230	3
SEQ ID NO: 3864	aagattttctctatgtgga	9561	9580	SEQ ID NO: 5193	ttccagaaaaactctctt	3936	3955	3
SEQ ID NO: 3865	gaaaaaacaggctgtgaag	9578	9597	SEQ ID NO: 5194	cttttacaactattttc	13021	13040	3
SEQ ID NO: 3866	tgaaagaattcttgaaaa	9590	9609	SEQ ID NO: 5195	tttttgagaatgaattcaa	10422	10441	3
SEQ ID NO: 3867	gaaggaaattctgaaaaac	9591	9610	SEQ ID NO: 5196	gtttgttgcgtglaaattca	11291	11310	3
SEQ ID NO: 3868	agctcagataagaaaaaac	9640	9659	SEQ ID NO: 5197	tttgataagtcacaaagct	9805	9824	3
SEQ ID NO: 3869	caaatctcttgacaggca	9720	9739	SEQ ID NO: 5198	gtctgagcagacattga	11688	11707	3
SEQ ID NO: 3870	ttgaaacaaaaataaattg	9788	9808	SEQ ID NO: 5199	gaactttgcactatgttcat	12762	12781	3
SEQ ID NO: 3871	aattcttgatacactgtt	9859	9878	SEQ ID NO: 5200	aacacatgaatcacaaatt	8938	8957	3
SEQ ID NO: 3872	ttcagctgttcaatgttga	9876	9896	SEQ ID NO: 5201	ttcaaaaagagctcaggaa	13207	13226	3
SEQ ID NO: 3873	aagtgtcttccattccact	9894	9913	SEQ ID NO: 5202	atggggaagtataagaact	4842	4861	3
SEQ ID NO: 3874	ttcagcatgtctatgttt	9950	9969	SEQ ID NO: 5203	agaaaaggacacactgtac	11080	11099	3
SEQ ID NO: 3875	ctgcacgtgggaattattc	10113	10132	SEQ ID NO: 5204	gttaagaaaatacagagcag	6440	6459	3
SEQ ID NO: 3876	tgaaatcaaatgtcgaact	10167	10186	SEQ ID NO: 5205	agattgaaggagacattca	7224	7243	3
SEQ ID NO: 3877	tattgtctcatcttcct	10201	10220	SEQ ID NO: 5206	aaggaaacataaactata	12889	12908	3
SEQ ID NO: 3878	gtgtctcatctcttct	10204	10223	SEQ ID NO: 5207	agaagaatactgcagaaca	12431	12450	3
SEQ ID NO: 3879	ctgtcttattgatgcactgc	10232	10251	SEQ ID NO: 5208	gcagttagactataagacga	13928	13947	3
SEQ ID NO: 3880	ccacagctctgtctctgag	10305	10324	SEQ ID NO: 5209	ctcagggatctgaaggtgg	8195	8214	3
SEQ ID NO: 3881	atttgaggaggtatgtcat	10330	10349	SEQ ID NO: 5210	atgaagtagacacacaaat	7161	7180	3
SEQ ID NO: 3882	atatggaagtgtcagtggc	10377	10396	SEQ ID NO: 5211	ttccacactccacgcata	10778	10797	3
SEQ ID NO: 3883	tggaataatcacaagcaaaa	10453	10472	SEQ ID NO: 5212	tttcaattcaattttcca	13023	13042	3
SEQ ID NO: 3884	aagtccaacacactgttct	10463	10482	SEQ ID NO: 5213	agacactgtgatactcact	12859	12878	3
SEQ ID NO: 3885	actgtctcttctccatgtg	10475	10494	SEQ ID NO: 5214	ccatgcagatgcagccagt	10924	10943	3

SEQ ID NO. 3888	ctctccatggaattaa	10482	10501	SEQ ID NO. 5215	taatcgagaggtalgaag	7148	7167	3
SEQ ID NO. 3887	atttctcaatgcgtactc	10512	10531	SEQ ID NO. 5216	gagttgagggtccgggaat	12242	12261	3
SEQ ID NO. 3888	tgacacacaagctagctt	10548	10567	SEQ ID NO. 5217	aagcgacactcaatcaaa	12036	12055	3
SEQ ID NO. 3889	ctcaccctctactttcc	10573	10592	SEQ ID NO. 5218	gggaactatgtagtgagg	10649	10668	3
SEQ ID NO. 3890	agcgcagggcacttccaa	10710	10729	SEQ ID NO. 5219	ttgggaagaagagcgagct	12289	12308	3
SEQ ID NO. 3891	tccaaatgtatgatact	10723	10742	SEQ ID NO. 5220	gatatacactaggaggagaa	12745	12764	3
SEQ ID NO. 3892	gagaacatacaagcaagc	10860	10879	SEQ ID NO. 5221	gcttggtttgccagttc	2467	2486	3
SEQ ID NO. 3893	atggcaaatgcagctctt	10897	10916	SEQ ID NO. 5222	aagaggtatttaaagccat	12960	12979	3
SEQ ID NO. 3894	tggaatgtcagctctgt	10898	10917	SEQ ID NO. 5223	caagaggtatttaaagcca	12959	12978	3
SEQ ID NO. 3895	gttcagggtccagtgcaag	10914	10933	SEQ ID NO. 5224	cttggggaggaggaacaa	14066	14085	3
SEQ ID NO. 3896	gttcagggtccagtgcaag	10915	10934	SEQ ID NO. 5225	acttggggaggaggaaca	14065	14084	3
SEQ ID NO. 3897	agttcttccatgatttcc	10940	10959	SEQ ID NO. 5226	gggaatcigtgaggaaact	12256	12275	3
SEQ ID NO. 3898	gdaacactgaagaaccag	10987	11006	SEQ ID NO. 5227	ctggatgaaccaccacca	11186	11205	3
SEQ ID NO. 3899	actaagaaccagaagatca	10994	11013	SEQ ID NO. 5228	tgatcaagaacctgttagt	13347	13366	3
SEQ ID NO. 3900	ctaagaaccagagatgact	10995	11014	SEQ ID NO. 5229	ctgatcaagaacctgttagt	13346	13365	3
SEQ ID NO. 3901	caagaatgacatgtgaaaa	11003	11022	SEQ ID NO. 5230	ttttcagaccaactctctg	13622	13641	3
SEQ ID NO. 3902	aaaaatgaagtcggatct	11018	11037	SEQ ID NO. 5231	gaatttgaaggtctgttt	9280	9299	3
SEQ ID NO. 3903	gattcattctgggtctctc	11032	11051	SEQ ID NO. 5232	gaaaacctatgctctaact	13166	13185	3
SEQ ID NO. 3904	agaaaaagcgaccactllga	11079	11098	SEQ ID NO. 5233	ctcaaaacctactgtctct	10486	10485	3
SEQ ID NO. 3905	aaggaacaccttaagttcct	11115	11134	SEQ ID NO. 5234	aggacacacaaatataactt	7572	7591	3
SEQ ID NO. 3906	ccagcattgttgaggagaca	11199	11218	SEQ ID NO. 5235	ctcaacaaagctaccacttg	12370	12389	3
SEQ ID NO. 3907	ctttgtgtacccaaataac	11239	11258	SEQ ID NO. 5236	gttttaaatgttgaaag	13148	13167	3
SEQ ID NO. 3908	ccatccctglaaaagtitt	11277	11296	SEQ ID NO. 5237	aaaagggtcatggaatgg	8893	8912	3
SEQ ID NO. 3909	gatcaaatcagttctt	11332	11351	SEQ ID NO. 5238	aagatagtcagtcigtalca	13334	13353	3
SEQ ID NO. 3910	aagaagctgagaacitcat	11432	11451	SEQ ID NO. 5239	atgagatcaacacaaactt	13110	13129	3
SEQ ID NO. 3911	ttgccctcaactaccacaa	11453	11472	SEQ ID NO. 5240	ttggtaagatgtactcaaa	12641	12660	3
SEQ ID NO. 3912	cttgattcccttttttag	11536	11555	SEQ ID NO. 5241	ctcaattttgattttcaag	8528	8547	3
SEQ ID NO. 3913	tcaagcttccaaaaaagtg	11591	11610	SEQ ID NO. 5242	cactcattgtattttgaa	12693	12712	3
SEQ ID NO. 3914	gtttcagatggcattgct	11608	11627	SEQ ID NO. 5243	agcagattatgttgaaca	11833	11852	3
SEQ ID NO. 3915	aattgcagtagccacaaga	11639	11658	SEQ ID NO. 5244	cttttcagcccaagcaatt	9231	9250	3
SEQ ID NO. 3916	ctgagcagacatttgagat	11691	11710	SEQ ID NO. 5245	atctgtaggagaactcag	12259	12278	3
SEQ ID NO. 3917	tgagcagacattgagatt	11692	11711	SEQ ID NO. 5246	aaactcagtaacaaactca	12256	12275	3
SEQ ID NO. 3918	ttgagattccctccattaa	11703	11722	SEQ ID NO. 5247	taacttcalaagttaa	13179	13198	3
SEQ ID NO. 3919	acttggagtgccagtgtga	11807	11826	SEQ ID NO. 5248	caatgtggagagagacagt	6504	6523	3
SEQ ID NO. 3920	caaatctgcgtacactcag	12004	12023	SEQ ID NO. 5249	ctgagacactcattcattg	11438	11457	3
SEQ ID NO. 3921	agccagcgttcaccgatac	12056	12075	SEQ ID NO. 5250	gtacacagtatagtgtgct	13286	13305	3
SEQ ID NO. 3922	cagcgttcaccgatacctca	12060	12079	SEQ ID NO. 5251	tggaacctgcaccaagctg	13960	13979	3
SEQ ID NO. 3923	ctccatctgcgtacacaga	12074	12093	SEQ ID NO. 5252	ctgatatacatcacagag	13711	13730	3
SEQ ID NO. 3924	atgaggaaacacagatcaa	12264	12283	SEQ ID NO. 5253	ttgagttgccaccatcat	11667	11686	3
SEQ ID NO. 3925	aggcagcttgcgttgtct	12300	12319	SEQ ID NO. 5254	agcaagcttctctggcct	3018	3037	3
SEQ ID NO. 3926	tgaagaacaagctgcccac	12327	12346	SEQ ID NO. 5255	ttgggagagacaaagttaa	6508	6527	3
SEQ ID NO. 3927	tatgatattgtacacaagt	12362	12381	SEQ ID NO. 5256	acttctcactatatttcat	12763	12782	3
SEQ ID NO. 3928	cattaggccaattatgat	12475	12494	SEQ ID NO. 5257	atcaacacactcttcaatg	13115	13134	3
SEQ ID NO. 3929	ttgactcaggaaggccaaag	12584	12603	SEQ ID NO. 5258	cttgtaagcagttactcaa	12640	12659	3
SEQ ID NO. 3930	gaacacctggatatacatc	12736	12755	SEQ ID NO. 5259	agtgattatacactcttctc	12865	12884	3
SEQ ID NO. 3931	tccttctgagttaaggaaa	12877	12896	SEQ ID NO. 5260	ttctgcacgtctcagga	13524	13543	3
SEQ ID NO. 3932	gccattcagctctcaaga	12974	12993	SEQ ID NO. 5261	cttccctctctgataatggc	5802	5821	3

SEQ ID NO: 3933	gtgctacgtaatdctcagg	13001	13020	SEQ ID NO: 5262	ctgtgaccacaaagctggcac	13964	13983	1	3
SEQ ID NO: 3934	agctgaagagatgaaatt	13065	13084	SEQ ID NO: 5263	aatttattcaaaacgagct	13200	13219	1	3
SEQ ID NO: 3935	aatttadctatctattaa	13080	13099	SEQ ID NO: 5264	ttaaaagaaatcttcaatt	13815	13834	1	3
SEQ ID NO: 3936	ttttaaaltgtgaagaa	13150	13169	SEQ ID NO: 5265	tctctctatgggaaaaaa	9566	9585	1	3
SEQ ID NO: 3937	taatcttcataagttcaat	13180	13199	SEQ ID NO: 5266	attgagattccctccatt	11702	11721	1	3
SEQ ID NO: 3938	tatatttgatccaagtata	13279	13298	SEQ ID NO: 5267	tataagcagaagcacatat	13937	13956	1	3
SEQ ID NO: 3939	tgaatatattgaacttga	13311	13330	SEQ ID NO: 5268	tcaaccttaattgttttca	8295	8314	1	3
SEQ ID NO: 3940	caattttgcacagaaata	13442	13461	SEQ ID NO: 5269	tattctctcttccaattg	13834	13853	1	3
SEQ ID NO: 3941	agaagattgcagagcttcc	13509	13528	SEQ ID NO: 5270	gaaatcttcaattattct	13821	13840	1	3
SEQ ID NO: 3942	gaagaaaaataatttctgat	13570	13589	SEQ ID NO: 5271	tatcagttcagataaacttc	7999	8018	1	3
SEQ ID NO: 3943	ttagcttgcattcaaaa	13680	13699	SEQ ID NO: 5272	tttgagaatgaatttcaa	10422	10441	1	3
SEQ ID NO: 3944	tcaaaactccacacattt	13693	13712	SEQ ID NO: 5273	aaattccttgacatgttga	7370	7389	1	3
SEQ ID NO: 3945	ttttttaaagaatcttc	13811	13830	SEQ ID NO: 5274	gaagtgctcagtgcaaaaa	10382	10401	1	3
SEQ ID NO: 3946	aggatctgagttatttgc	14043	14062	SEQ ID NO: 5275	gcaaggggtcacgtgttct	7864	7883	1	3
SEQ ID NO: 3947	tttgctaaacttgggggag	14057	14076	SEQ ID NO: 5276	ctccccaggaccttcaaa	9842	9861	1	3

Table 11. Selected palindromic sequences from human glucose-6-phosphatase

	Source	Start Index	End Index		Match	Start Index	End Index	#
SEQ ID NO: 5291	tccatcttcagggaagctgt	222	241	SEQ ID NO: 5369	acagactcttcagatgga	1340	1358	16
SEQ ID NO: 5292	ccatcttcagggaagctgtg	223	242	SEQ ID NO: 5370	cacagactcttcagatgg	1339	1358	16
SEQ ID NO: 5293	cctctggccactccatggg	417	436	SEQ ID NO: 5371	gccatttggaggccagagg	1492	1511	16
SEQ ID NO: 5294	ctcttcaggccatccatgggc	418	437	SEQ ID NO: 5372	cccatcttggaggccagag	1491	1510	16
SEQ ID NO: 5295	tgaatgttcatttttggt	521	540	SEQ ID NO: 5373	cccatcatattatcatcaa	2945	2964	16
SEQ ID NO: 5296	tcagtaaatgggggaccgac	1886	1905	SEQ ID NO: 5374	ctggctctcgaaactctga	2731	2750	16
SEQ ID NO: 5297	ttttactgtctcacaatgt	1956	1975	SEQ ID NO: 5375	tcactcttgaagaagaaaa	2983	3002	16
SEQ ID NO: 5298	ggaggtgcgaaggaatga	50	69	SEQ ID NO: 5376	catgtctcagcgtctcca	2620	2639	16
SEQ ID NO: 5299	ggaggtgcgaaggaatga	51	70	SEQ ID NO: 5377	ctcatgtctcagcgtctcc	2619	2638	16
SEQ ID NO: 5300	gggaagaataaagccgacc	487	506	SEQ ID NO: 5378	tcctctctggctgtatttcc	1295	1314	15
SEQ ID NO: 5301	ttcttcactcagaatgtgt	598	617	SEQ ID NO: 5379	taacatcttgaagaagaaa	2982	3001	15
SEQ ID NO: 5302	ctcttcagcatccacagac	651	670	SEQ ID NO: 5380	cttggacatctggagaaga	773	792	15
SEQ ID NO: 5303	tgagctctggagaagagcc	776	795	SEQ ID NO: 5381	ggggctggctctcaactcca	884	903	15
SEQ ID NO: 5304	agcctctccagaacccgtg	848	867	SEQ ID NO: 5382	cagactctctccatggct	2107	2126	15
SEQ ID NO: 5305	ggcctctgggggtcgtctca	878	897	SEQ ID NO: 5383	tgaggccaacgagccggcc	2801	2820	15
SEQ ID NO: 5306	ggagctcactccactcgaa	1439	1458	SEQ ID NO: 5384	ttccagtgtagggccagctc	1676	1695	15
SEQ ID NO: 5307	agctaagtgaagcatatgg	1572	1591	SEQ ID NO: 5385	ctcagcctctccatagtag	2626	2645	15
SEQ ID NO: 5308	tccttcgaataagcatatgga	1573	1592	SEQ ID NO: 5386	ctcagcctctccatagtag	2626	2644	15
SEQ ID NO: 5309	ctaaatgtgcttaattata	1854	1873	SEQ ID NO: 5387	tatatltttagaattatg	2683	2702	15
SEQ ID NO: 5310	ctctctctctctctctctct	2509	2528	SEQ ID NO: 5388	gaaaaatatatattgtcag	2996	3015	15
SEQ ID NO: 5311	caatcaccacacagccgtg	0	19	SEQ ID NO: 5389	cgagtaagggtccactatg	812	831	15
SEQ ID NO: 5312	agcctggaaataactgtcaag	12	31	SEQ ID NO: 5390	cttggaattctggaatggct	1987	2006	14
SEQ ID NO: 5313	gtgtccatctctcaggagact	220	239	SEQ ID NO: 5391	agctcactccactcaggaac	1440	1459	14
SEQ ID NO: 5314	gtgtgggttggttgtaactg	326	345	SEQ ID NO: 5392	cgactctccaccactcca	2425	2444	14
SEQ ID NO: 5315	taactgtgtagactgtggcag	392	411	SEQ ID NO: 5393	ctggagaagccagagaggt	782	801	14
SEQ ID NO: 5316	gctgtctgagcaacttcca	638	657	SEQ ID NO: 5394	tgatgtgtctcttcgcagc	1474	1493	14
SEQ ID NO: 5317	tcagacatcataatgcagac	666	685	SEQ ID NO: 5395	ctgggttgactctcactcgt	758	777	14
SEQ ID NO: 5318	gggtgtgtagactcctgtgg	760	779	SEQ ID NO: 5396	cccatgtgacacacacacc	823	842	14
SEQ ID NO: 5319	gtgtgtgtagactcctctgga	761	780	SEQ ID NO: 5397	tcacattgacaccacacac	822	841	14
SEQ ID NO: 5320	gtgtgagacctcctgtggac	762	781	SEQ ID NO: 5398	gtccacattgacacacac	821	840	14
SEQ ID NO: 5321	gaacctctctggaactcgtg	767	786	SEQ ID NO: 5399	cagatattgtcactcaggtc	2014	2033	14
SEQ ID NO: 5322	ctctctggcgacgcctctggc	862	881	SEQ ID NO: 5400	cgctccacgcagcagccagg	1687	1706	14
SEQ ID NO: 5323	ctgggcacgcctctttggcc	863	882	SEQ ID NO: 5401	ggccagctctcacagccag	1686	1705	14
SEQ ID NO: 5324	ctggctctcactcctctgtg	1028	1047	SEQ ID NO: 5402	acaaaaggcaagctctcag	1663	1682	14
SEQ ID NO: 5325	gtatgagctgtgtagtggccct	1056	1075	SEQ ID NO: 5403	agagctctcagatctctct	2229	2248	14
SEQ ID NO: 5326	gggcacttggatttggag	1217	1236	SEQ ID NO: 5404	ctcccactgtgacagcccca	1446	1465	14
SEQ ID NO: 5327	gaattaaataacagtgatgg	1267	1286	SEQ ID NO: 5405	ggcacaacaggagacacatc	2311	2330	14
SEQ ID NO: 5328	ttgtctcagaatgttggtt	1598	1617	SEQ ID NO: 5406	aacatctctgctcataaca	2967	2986	14
SEQ ID NO: 5329	aggagcctctgaactgtata	1764	1783	SEQ ID NO: 5407	tatacatattacatctct	2063	2082	14
SEQ ID NO: 5330	taaaagtcttaattatat	1855	1874	SEQ ID NO: 5408	atatatgtcagatttata	3003	3022	14
SEQ ID NO: 5331	aaaaatgcaagggaggggc	2215	2234	SEQ ID NO: 5409	ggccctctctgcgttgtt	2817	2836	14
SEQ ID NO: 5332	taaaagaaaagatcaacat	2330	2349	SEQ ID NO: 5410	algtgcgatttatttaa	3007	3026	14
SEQ ID NO: 5333	taaaagtctctctcttttt	2345	2364	SEQ ID NO: 5411	aaaaagaaaatatatatgt	2992	3011	14
SEQ ID NO: 5334	ttctacgtctctctccoca	197	216	SEQ ID NO: 5412	ggggccagccgcacaaagaa	1116	1135	14

SEQ ID NO: 5335	gggtagctgtgattggag	257	276	SEQ ID NO: 5413	ctcccactggaacagccca	1446	1465	13
SEQ ID NO: 5336	gctgtgattggagactggc	263	282	SEQ ID NO: 5414	gccatgccatgggcacagc	423	442	13
SEQ ID NO: 5337	cacttcctgtgcccttgata	358	377	SEQ ID NO: 5415	tatccaccaggctggagtg	2548	2567	13
SEQ ID NO: 5338	acatctactctttccaact	464	483	SEQ ID NO: 5416	agatgggtatttcacatgt	2705	2724	13
SEQ ID NO: 5339	ctactctttccactttcca	468	487	SEQ ID NO: 5417	gaatactctcacaaagtag	1419	1438	13
SEQ ID NO: 5340	cagataaaaggccgacctag	492	511	SEQ ID NO: 5418	ctgttttccaactcactc	2828	2847	13
SEQ ID NO: 5341	gtgcagctgaatgtctgt	553	572	SEQ ID NO: 5419	acagaaactttcagccaca	644	663	13
SEQ ID NO: 5342	atgtctgtctgtcagcaat	564	583	SEQ ID NO: 5420	attcaggctatagctgacat	2038	2057	13
SEQ ID NO: 5343	ctgtgcacgaatctactctg	572	591	SEQ ID NO: 5421	caagggtctagattacag	2779	2798	13
SEQ ID NO: 5344	atcaagttgtgtcggagt	606	625	SEQ ID NO: 5422	actctgcagcctaagtgat	2742	2761	13
SEQ ID NO: 5345	cagaaactttcagccacat	645	664	SEQ ID NO: 5423	atgtttcaattaggctctg	2185	2204	13
SEQ ID NO: 5346	actttcagccacatccaca	650	669	SEQ ID NO: 5424	gtggcgatcatgcaagt	1818	1837	13
SEQ ID NO: 5347	atgccagcctcaagaataa	678	697	SEQ ID NO: 5425	tatttttttactgtgcat	1950	1969	13
SEQ ID NO: 5348	agaaatattttcattacc	690	709	SEQ ID NO: 5426	gtaaatatgactcctttc	2283	2302	13
SEQ ID NO: 5349	gaatatattttcattacc	691	710	SEQ ID NO: 5427	ggtaaatatgactcctttc	2282	2301	13
SEQ ID NO: 5350	tgctgctcaaggactggg	744	763	SEQ ID NO: 5428	cccaaggcaaccaagagca	2306	2325	13
SEQ ID NO: 5351	ctctgtgactctggagaaa	772	791	SEQ ID NO: 5429	tttcalcatgttggccagg	2713	2732	13
SEQ ID NO: 5352	ggagaaagcccagagggg	784	803	SEQ ID NO: 5430	ccacgcgcacggggccctcc	2805	2824	13
SEQ ID NO: 5353	ttgaacccccatcccaag	1004	1023	SEQ ID NO: 5431	tttgatacttctggcctcaa	2405	2424	13
SEQ ID NO: 5354	cagatggagggtgccatctc	1351	1370	SEQ ID NO: 5432	gatatcgagagattttctg	2847	2866	13
SEQ ID NO: 5355	ggagctcactcccacttga	1438	1457	SEQ ID NO: 5433	tccacctgccttggccctcc	2760	2779	13
SEQ ID NO: 5356	ttgggtatgttttga	1553	1572	SEQ ID NO: 5434	tttctctatcccaagccaa	2297	2316	13
SEQ ID NO: 5357	gaagtgggtgtttcttga	1606	1625	SEQ ID NO: 5435	tccacccactgggactctc	2131	2150	13
SEQ ID NO: 5358	aaaagaaggctgcctaaagg	1785	1804	SEQ ID NO: 5436	ccttgcctgtcttttttt	2503	2522	13
SEQ ID NO: 5359	aaagaaggctgcctaaagg	1786	1805	SEQ ID NO: 5437	tccttgcctgtctttttt	2502	2521	13
SEQ ID NO: 5360	aaagaaggctgcctaaagg	1787	1806	SEQ ID NO: 5438	ctcctgtcctgtcttttt	2501	2520	13
SEQ ID NO: 5361	agaaggctgcctaaagg	1788	1807	SEQ ID NO: 5439	cctctgtcctgtcttttt	2500	2519	13
SEQ ID NO: 5362	atttctgtgatttttga	1982	2001	SEQ ID NO: 5440	tcaattaggctctgaat	2189	2208	13
SEQ ID NO: 5363	tccttataagcccagctct	2081	2100	SEQ ID NO: 5441	agagcacattcttaagaga	2319	2338	13
SEQ ID NO: 5364	ataagcccagctctgtctt	2086	2105	SEQ ID NO: 5442	aaagctgaagcctatttat	2889	2908	13
SEQ ID NO: 5365	ggccaggaattctctctca	2231	2250	SEQ ID NO: 5443	gagccaccgcaccgggcc	2801	2820	13
SEQ ID NO: 5366	gccaactcctcctgtcctg	2493	2512	SEQ ID NO: 5444	caggctggagtggaaggc	2555	2574	13
SEQ ID NO: 5367	ttttttttttttttgag	2519	2538	SEQ ID NO: 5445	ctcataacattcttga	2977	2996	13
SEQ ID NO: 5368	cggcgctgcaccaccatgc	2652	2671	SEQ ID NO: 5446	gcattgagccaccgcaccg	2798	2817	13

Table 12. Selected palindromic sequences from rat glucose-6-phosphatase

	Source	Start Index	End Index		Match	Start Index	End Index	#B
SEQ ID NO: 5447	ctgactattacagcaacag	301	320	SEQ ID NO: 5471	ctgaggctgaaacttcag	598	617	16
SEQ ID NO: 5448	ctctgggggttggggctgg	831	850	SEQ ID NO: 5472	ccagcatgtaccgcaagag	859	878	16
SEQ ID NO: 5449	tgcaaggagaaactgcgca	879	898	SEQ ID NO: 5473	tcgacgcgtcccctttgca	1019	1038	16
SEQ ID NO: 5450	ctcggggccatgccatggg	376	395	SEQ ID NO: 5474	cccagtgfggggcccagagg	1171	1190	16
SEQ ID NO: 5451	ttgagcaaacccatattgcaa	1478	1497	SEQ ID NO: 5475	ttgcagagtggtcttctcaa	2057	2076	15
SEQ ID NO: 5452	cagcttctgagggtaccac	2	21	SEQ ID NO: 5476	ttggtlctglgatcgctg	123	142	14
SEQ ID NO: 5453	gggtaccaaggagggaagat	13	32	SEQ ID NO: 5477	atccagtcgactcgctacc	66	85	14
SEQ ID NO: 5454	ctccacgactttgggatcc	51	70	SEQ ID NO: 5478	ggatcgggaggagggggag	1448	1467	14
SEQ ID NO: 5455	caggactgggtttgcttgg	108	127	SEQ ID NO: 5479	ccaagcccgcactgtgcctg	2018	2037	14
SEQ ID NO: 5456	cttctatgtctcttctccc	155	174	SEQ ID NO: 5480	gggacagacacacagaag	1076	1095	14
SEQ ID NO: 5457	ttctatgtctcttctccca	156	175	SEQ ID NO: 5481	tgggacagacacacaagaa	1075	1094	14
SEQ ID NO: 5458	tggtccacattcaagaga	177	196	SEQ ID NO: 5482	ctcctaatagatagacca	1549	1568	14
SEQ ID NO: 5459	gcctctgatataaaacagtt	325	344	SEQ ID NO: 5483	aaactdgagactttgggca	1868	1887	14
SEQ ID NO: 5460	agcccggtctctgggacag	1064	1083	SEQ ID NO: 5484	ctgtccccagcctgggct	2034	2053	14
SEQ ID NO: 5461	agctctgcacacaagtcag	1111	1130	SEQ ID NO: 5485	ctgaatggatgatgtgact	1659	1678	14
SEQ ID NO: 5462	aaaaagggtgaatttttaa	1237	1256	SEQ ID NO: 5486	tttataaaacgacatttt	2201	2220	14
SEQ ID NO: 5463	acactctcaataatgatag	1545	1564	SEQ ID NO: 5487	ctatgaatgatgctgtgt	2121	2140	14
SEQ ID NO: 5464	aaagaatgaactgtgtcca	37	56	SEQ ID NO: 5488	tggaactctctgtgacttt	724	743	13
SEQ ID NO: 5465	cttgggattccagtcgact	59	78	SEQ ID NO: 5489	agtcagcggccgtgcaaa	1124	1143	13
SEQ ID NO: 5466	gtgatcgctgacctcagga	132	151	SEQ ID NO: 5490	ccctctctccaaaggctac	1911	1930	13
SEQ ID NO: 5467	ggaacgctcttctgtctc	148	167	SEQ ID NO: 5491	aggactctactctcttcc	1748	1767	13
SEQ ID NO: 5468	gactgtgggcatacaatctc	194	213	SEQ ID NO: 5492	ggactggaccaggaggctc	357	376	13
SEQ ID NO: 5469	ggacactgactattacagc	296	315	SEQ ID NO: 5493	gctgaacgtctgtgttcc	518	537	13
SEQ ID NO: 5470	aagcccccgctccagattg	966	985	SEQ ID NO: 5494	caattgttctgctgtctt	1833	1852	13

Table 13. Selected palindromic sequences from human B-catenin

	Source	Start Index	End Index		Match	Start Index	End Index	#B
SEQ ID NO: 5495	agcagcttcagctccccgcc	70	89	SEQ ID NO: 5542	ggcgacatgacgactgct	2152	2171	15
SEQ ID NO: 5496	ccattctggtgcacactcc	304	323	SEQ ID NO: 5543	ggatggaccocatgatgg	2387	2406	15
SEQ ID NO: 5497	tccttctctgagtgtaaa	328	347	SEQ ID NO: 5544	ttattacatcaagaagga	985	1004	15
SEQ ID NO: 5498	ctctgagtggaaggaat	334	353	SEQ ID NO: 5545	attgatctgacctacgaga	791	810	15
SEQ ID NO: 5499	cagagggatcagcagctcta	473	492	SEQ ID NO: 5546	agctgcgggggtctctgt	2037	2056	15
SEQ ID NO: 5500	ctaaatgacgaggaccagg	677	696	SEQ ID NO: 5547	tcctgaaatcatctcttag	2539	2558	15
SEQ ID NO: 5501	taaatgacgaggaccagg	678	697	SEQ ID NO: 5548	acctgtaaatcatctctta	2538	2557	15
SEQ ID NO: 5502	gtctctgtatggtggagca	383	402	SEQ ID NO: 5549	gttcgcgaalgtctggagc	2176	2195	24
SEQ ID NO: 5503	cccagcgccgtacgtccat	1839	1858	SEQ ID NO: 5550	atgggtgctgccagctcgg	2451	2470	24
SEQ ID NO: 5504	ccccctggaggtattggaa	143	162	SEQ ID NO: 5551	ttcacatctagctcgggag	1929	1948	14
SEQ ID NO: 5505	gggtatttgaaagtatacca	151	170	SEQ ID NO: 5552	tggttaagctcttacacc	1680	1699	14
SEQ ID NO: 5506	gctgttagtcaactgcgcag	260	279	SEQ ID NO: 5553	gctgcctccagctgacagc	2494	2513	14
SEQ ID NO: 5507	gtctctgtatggtggagca	383	402	SEQ ID NO: 5554	gttcgctctacatctggac	1652	1671	14
SEQ ID NO: 5508	ctctgtatgagtggaaca	384	403	SEQ ID NO: 5555	gttcgcgaatgtctgagga	2175	2194	14
SEQ ID NO: 5509	gtatgcgaatgactcagct	454	473	SEQ ID NO: 5556	agctgcctgggttgatctg	2517	2536	14
SEQ ID NO: 5510	gtccacgctgtgtgtgaac	563	582	SEQ ID NO: 5557	gttcgctctacatctggac	1652	1671	14
SEQ ID NO: 5511	tatcaaatgatgcagagca	623	642	SEQ ID NO: 5558	gttcgtgcacatcaggata	1820	1839	14
SEQ ID NO: 5512	tatggtccatcagctctct	718	737	SEQ ID NO: 5559	agaagaacaagctcatcata	1126	1145	14
SEQ ID NO: 5513	ccccctgggaaatgctcgg	915	934	SEQ ID NO: 5560	ccaagaatgactgcaggg	2029	2048	14
SEQ ID NO: 5514	agctttaggacttcacctg	1291	1310	SEQ ID NO: 5561	caggtagacagcaatcagct	2502	2521	14
SEQ ID NO: 5515	ggaaatcttcagatgtcgc	1356	1375	SEQ ID NO: 5562	gcagctgcgtttgtttgcc	2162	2181	14
SEQ ID NO: 5516	gtcctctgggctgggtgac	1549	1568	SEQ ID NO: 5563	gtcctctgcacagcgcgaca	1605	1624	14
SEQ ID NO: 5517	ccagactctctgcacagag	2107	2126	SEQ ID NO: 5564	ctctaggaatgaagggtgtg	2134	2153	14
SEQ ID NO: 5518	ccagacagaaaagcggctg	245	264	SEQ ID NO: 5565	cagctcgtgtgacccgtg	828	847	23
SEQ ID NO: 5519	cagcagcgtgtggccggcc	4	23	SEQ ID NO: 5566	ggccaccacccctgggtgct	2420	2439	13
SEQ ID NO: 5520	aggtctgaggagcagctct	60	79	SEQ ID NO: 5567	gaagaggatgggataacct	359	378	13
SEQ ID NO: 5521	actgttttgaaaatccagc	174	193	SEQ ID NO: 5568	gctgatattgatggacagct	437	456	13
SEQ ID NO: 5522	gtatttgatggagtggga	213	232	SEQ ID NO: 5569	ccaggtagacagcaatcag	2500	2519	13
SEQ ID NO: 5523	ccagacagaaaagcggctg	245	264	SEQ ID NO: 5570	cagcaaacacttaccctgg	275	294	13
SEQ ID NO: 5524	acagctcctctctgagtg	323	342	SEQ ID NO: 5571	caclgagctgctccactgt	1579	1598	13
SEQ ID NO: 5525	tggaactcctccaaagctct	369	388	SEQ ID NO: 5572	aggactaaatcaccattcca	1972	1991	13
SEQ ID NO: 5526	taagaacaacagctgctgat	424	443	SEQ ID NO: 5573	atcagctggcctgtgttga	2514	2533	13
SEQ ID NO: 5527	agctcagaggggtacagct	469	488	SEQ ID NO: 5574	agctgtgtggaatgcaagct	1276	1295	13
SEQ ID NO: 5528	gcagtcagatcccatctac	516	535	SEQ ID NO: 5575	gtagaagctgggtgaagct	1271	1290	13
SEQ ID NO: 5529	ccacacgtgcacactccga	645	664	SEQ ID NO: 5576	tcagatgatataaagtgg	1430	1449	13
SEQ ID NO: 5530	ccacagctgcaatccctgaa	646	665	SEQ ID NO: 5577	tcagatgatataaagtgg	1429	1448	13
SEQ ID NO: 5531	ggacactgtcaaacactttc	846	865	SEQ ID NO: 5578	gaaacttgcctcttgc	1743	1762	13
SEQ ID NO: 5532	ctccacaacacctttattac	974	993	SEQ ID NO: 5579	gtaaatcatctcttaggag	2542	2561	13
SEQ ID NO: 5533	cagagtgtctgaagtgctga	1222	1241	SEQ ID NO: 5580	agctgcaggggtcctctg	2037	2056	13
SEQ ID NO: 5534	ggactcagctcaggactttc	1347	1366	SEQ ID NO: 5581	gaaacttgcctcttgc	1743	1762	13
SEQ ID NO: 5535	tgatataaagtgtgtcacc	1435	1454	SEQ ID NO: 5582	ggtgcacagggaagacata	1562	1581	13
SEQ ID NO: 5536	cccagcgccgtacgtccat	1839	1858	SEQ ID NO: 5583	atgcccagaggtcctgtg	2370	2389	13
SEQ ID NO: 5537	gtccatctgtggtggacacg	1852	1871	SEQ ID NO: 5584	ctgtgaactgtctcaggag	2053	2072	13
SEQ ID NO: 5538	gtgaccggagcccttcac	1915	1934	SEQ ID NO: 5585	gtgaactgtctcaggacaa	2055	2074	13
SEQ ID NO: 5539	ttgattacagaggaataaa	1962	1981	SEQ ID NO: 5586	ttaggagtaacaatacaa	2553	2572	13
SEQ ID NO: 5540	gaagctatgaaactcaggg	2084	2103	SEQ ID NO: 5587	ctctgcagagagttacttc	2114	2133	13
SEQ ID NO: 5541	tcagaacagagccaatggc	2247	2266	SEQ ID NO: 5588	gccacacccctgggtctga	2421	2440	13

Table 14. Selected palindromic sequences from human hepatitis C virus (HCV)

	Source	Start Index	End Index		Match	Start Index	End Index	#
SEQ ID NO:	5589cagcactcgggtgctcggt	5314	5333	SEQ ID NO:	6135taccatcaccacgctcgtc	6196	6216	19
SEQ ID NO:	5590aacctgcgcggatgcgcgg	1682	1701	SEQ ID NO:	6136ccgggcagccgggtcgaatt	8202	8221	18
SEQ ID NO:	5591cgctcgtggtgacgcgtca	1049	1068	SEQ ID NO:	6137cgagcgcagccgcgcgcg	6151	6170	17
SEQ ID NO:	5592ctccgcgtaccacaaagcg	1352	1371	SEQ ID NO:	6138cggcatgtggggcccgagg	6053	6072	17
SEQ ID NO:	5593tgtaacatcgggggggtcg	2048	2067	SEQ ID NO:	6139cgaccctccacattaca	6871	6890	17
SEQ ID NO:	5594gtaacatcgggggggtcgg	2049	2068	SEQ ID NO:	6140ccgacccctccacattac	6870	6889	17
SEQ ID NO:	5595cagccaccaacagggcgga	5556	5575	SEQ ID NO:	6141tcggcgtgttcgtgtcgt	9254	9273	17
SEQ ID NO:	5596ctcaccacccagaaacccc	5744	5763	SEQ ID NO:	6142gggtgtgcacgggtgtgag	6291	6310	17
SEQ ID NO:	5597ccagccttaccatcaccca	6189	6208	SEQ ID NO:	6143ggggcgtgtgatgtcgtg	5832	5851	17
SEQ ID NO:	5598ctacgcgcgtgttcggcgtc	6249	6268	SEQ ID NO:	6144gagcccggaacggcagtag	6830	6849	17
SEQ ID NO:	5599tacgcgcgtgttcggcgtc	6250	6269	SEQ ID NO:	6145cgagcccggaacggcagtag	6829	6848	17
SEQ ID NO:	5600gagttcctgtgtaaaagcct	8216	8235	SEQ ID NO:	6146aggctatgatcaggtaactc	8634	8653	17
SEQ ID NO:	5601atggcgggggaactgggcta	1430	1449	SEQ ID NO:	6147tagccgattttactccat	9019	9038	26
SEQ ID NO:	5602aaccaaacagctaacacaac	370	389	SEQ ID NO:	6148gttggcgcatacctaggtt	4115	4134	16
SEQ ID NO:	5603gggtgctcagatcgtgtgtg	419	438	SEQ ID NO:	6149cacagcccgctacaccacc	5734	5753	16
SEQ ID NO:	5604ccttgccgcccctcatgcca	584	603	SEQ ID NO:	6150tgccaacgtgggtacaagg	6374	6393	16
SEQ ID NO:	5605taccgcgcgcgcgcgcgcag	1265	1284	SEQ ID NO:	6151ctgacgactacgtcgcgta	8465	8484	16
SEQ ID NO:	5606gggcgcgcgtgccgcgtcca	1508	1527	SEQ ID NO:	6152tgagcagcagccgcgtgcc	4759	4778	16
SEQ ID NO:	5607ctgcaatgactcctccag	1624	1643	SEQ ID NO:	6153ctgggtggccctcaatgcag	2594	2613	16
SEQ ID NO:	5608aacgcgtcgtctcggcaac	1897	1916	SEQ ID NO:	6154gttgcgcatacctagggt	4115	4134	16
SEQ ID NO:	5609gtgcggggcccccgcgtgt	2032	2051	SEQ ID NO:	6155aacaccagggccccctgcac	6537	6556	16
SEQ ID NO:	5610atgtggggggcgltggaaga	2238	2257	SEQ ID NO:	6156tgctcaatgtcctacacat	7610	7629	16
SEQ ID NO:	5611ggagagcgtgttcaacttgg	2288	2307	SEQ ID NO:	6157ccaagctaaactcactcc	9207	9226	16
SEQ ID NO:	5612cgtcgtgtgccggagcgca	2613	2632	SEQ ID NO:	6158tgagcggccgaaccggcagc	6827	6846	16
SEQ ID NO:	5613gtctggcattattgacact	2817	2836	SEQ ID NO:	6159aaggctacacttgcagacg	7763	7782	16
SEQ ID NO:	5614ctttgatataccaaaact	2997	3016	SEQ ID NO:	6160agtctgatgaatggaaga	5454	5473	16
SEQ ID NO:	5615ctcttgattgcacatactg	3014	3033	SEQ ID NO:	6161cgagcaattcaagcagaag	5518	5537	16
SEQ ID NO:	5616ggcggcggtlgtgggacatca	3314	3333	SEQ ID NO:	6162tgatcagccatgcgcgcgc	7641	7660	16
SEQ ID NO:	5617gggacatcatctgtgggct	3324	3343	SEQ ID NO:	6163aggcgggtggattttgtccc	3915	3934	16
SEQ ID NO:	5618gggcgtctctcgggcgcgt	3874	3893	SEQ ID NO:	6164agcggcgccggcgaccggcc	7439	7458	16
SEQ ID NO:	5619ggcggtctccggggccgcgt	3875	3894	SEQ ID NO:	6165cagcgccagcgccagccgc	7438	7457	16
SEQ ID NO:	5620gcgtctctcgggcgcgtgt	3876	3895	SEQ ID NO:	6166acaggtgcccctgatacgc	7631	7650	16
SEQ ID NO:	5621gtcccggtgtcttcacagac	3961	3980	SEQ ID NO:	6167gtcttggaagaaccggcag	7252	7271	16
SEQ ID NO:	5622catcaggaactgggglaagg	4174	4193	SEQ ID NO:	6168ccttcagagccgctgagt	8155	8174	16
SEQ ID NO:	5623ccgacgggtggttgcctcgg	4245	4264	SEQ ID NO:	6169ccgggggaacggccctcgg	4853	4872	16
SEQ ID NO:	5624gggggggaagcgcaactcatt	4501	4520	SEQ ID NO:	6170aatgttgtgacttggcccc	8334	8353	16
SEQ ID NO:	5625ccgagcaattcaagcagaa	5517	5536	SEQ ID NO:	6171ttcgattgtccatactcgg	3015	3034	16
SEQ ID NO:	5626agatgaaggcaaaaggcgltc	7821	7840	SEQ ID NO:	6172gacgaccttctgtttatct	8564	8583	16
SEQ ID NO:	5627ccccctaggggcgctgcga	767	786	SEQ ID NO:	6173ggccggcgccccccggggg	3674	3693	35
SEQ ID NO:	5628ctcccggtcgtgtggggc	646	665	SEQ ID NO:	6174gcccccccttgagggggag	7519	7538	25
SEQ ID NO:	5629ttccgtctcgtcggcgccc	750	769	SEQ ID NO:	6175gggcaaaaggacgtcgggaa	7923	7942	25
SEQ ID NO:	5630ccccctaggggcgctgcga	767	786	SEQ ID NO:	6176ggcgggggcccaactgggg	1383	1402	25
SEQ ID NO:	5631gcccccgccgcatgcgaca	1222	1241	SEQ ID NO:	6177gtccagagggggcgagggc	9147	9166	25
SEQ ID NO:	5632aggagcagccgggtctcttc	178	197	SEQ ID NO:	6178gaaaaggagcggttgtcct	7341	7360	15
SEQ ID NO:	5633ggagcagccgggtctcttct	179	198	SEQ ID NO:	6179gaaaaaggagcggttgtcc	7340	7359	15

SEQ ID NO: 5634	aaacacaaacaglaacacca	368	387	SEQ ID NO: 6180	gggttttttttttt	9443	9462	15
SEQ ID NO: 5635	caaccgcgcgcacacaggac	385	404	SEQ ID NO: 6181	glctcgacccgctclgtg	4100	4119	15
SEQ ID NO: 5636	cgggtgtcagatggtgtg	418	437	SEQ ID NO: 6182	accatgtgacgacgacg	4754	4773	15
SEQ ID NO: 5637	actcgtgcgcgcgcagg	444	463	SEQ ID NO: 6183	ccccgcgcacgctcaggt	1267	1286	15
SEQ ID NO: 5638	gcgcgcgcgcgcgcgcgc	450	469	SEQ ID NO: 6184	ctggcgccgcctgcgcgcgc	3164	3183	15
SEQ ID NO: 5639	gggccccagggtgggtgtg	480	479	SEQ ID NO: 6185	cacagcctgtctgtgtccc	9296	9315	15
SEQ ID NO: 5640	gtgggggcgcgcgcgcgc	657	676	SEQ ID NO: 6186	gggtgggtgacgcgcgcgc	5763	5802	15
SEQ ID NO: 5641	gtgggcgcgcgcgcgcgc	658	677	SEQ ID NO: 6187	gggtgggtgacgcgcgcgc	5782	5801	15
SEQ ID NO: 5642	ggggccccacgcgcgcgc	659	678	SEQ ID NO: 6188	gggggtgggtgacgcgcgc	5781	5800	15
SEQ ID NO: 5643	ccctcacatgcgcgcgcgc	715	734	SEQ ID NO: 6189	gggggggcgcgcgcgcgc	3774	3793	15
SEQ ID NO: 5644	ccatcgcgcgcgcgcgcgc	718	737	SEQ ID NO: 6190	gtgcgcgcgcgcgcgcgc	6020	6039	15
SEQ ID NO: 5645	tcgcgcgcgcgcgcgcgc	751	770	SEQ ID NO: 6191	ggggcacaaggagctgcgc	7922	7941	15
SEQ ID NO: 5646	ggcgctgcgcgcgcgcgc	776	795	SEQ ID NO: 6192	ccagccacacgtgtgcgc	6110	5129	15
SEQ ID NO: 5647	tcgatgcgcgcgcgcgcgc	943	962	SEQ ID NO: 6193	gcagccacacgtgtgcgc	6498	6517	15
SEQ ID NO: 5648	gtgcgcgcgcgcgcgcgc	1019	1038	SEQ ID NO: 6194	ccctcacacgcgcgcgcgc	1495	1514	15
SEQ ID NO: 5649	gtgcgcgcgcgcgcgcgc	1020	1039	SEQ ID NO: 6195	ccctcacacgcgcgcgcgc	1494	1513	15
SEQ ID NO: 5650	ggcctcggtgtgcgcgcgc	1021	1040	SEQ ID NO: 6196	accctcacacgcgcgcgc	1493	1512	15
SEQ ID NO: 5651	aggaaatgcgcgcgcgcgc	1085	1104	SEQ ID NO: 6197	gggcacgtgcgcgcgcgc	4323	4342	15
SEQ ID NO: 5652	ccccacacgcgcgcgcgc	1098	1117	SEQ ID NO: 6198	cgatgtccacagatgtgga	8092	8111	15
SEQ ID NO: 5653	atacgcacacacgcgcgc	1112	1131	SEQ ID NO: 6199	aatcactgctgacgcgc	4578	4597	15
SEQ ID NO: 5654	attgtcgtgtgcgcgcgc	1128	1147	SEQ ID NO: 6200	ggcgccactgtgcgcgcgc	9184	9183	15
SEQ ID NO: 5655	cccttcctgcgcgcgcgcgc	1215	1234	SEQ ID NO: 6201	ggcaacgtgtgcgcgcgc	6374	6393	15
SEQ ID NO: 5656	gaacccgcgcgcgcgcgc	1266	1285	SEQ ID NO: 6202	ctgcgcgcgcgcgcgcgc	6340	6359	15
SEQ ID NO: 5657	ggcctcgtgtgtgcgcgc	1331	1350	SEQ ID NO: 6203	actgcgcgcgcgcgcgcgc	6046	6065	15
SEQ ID NO: 5658	ggcgtctcaggaatcgcgc	1558	1577	SEQ ID NO: 6204	clgtatgcgcgcgcgcgc	5838	5857	15
SEQ ID NO: 5659	ctgaactgcgcgcgcgcgc	1619	1638	SEQ ID NO: 6205	gggcacacgtgcgcgcgc	2313	2332	15
SEQ ID NO: 5660	agactggtgtgtgtgcgc	1641	1660	SEQ ID NO: 6206	ggcgccgcgcgcgcgcgc	8609	8628	15
SEQ ID NO: 5661	tgctgcgcgcgcgcgcgc	1685	1704	SEQ ID NO: 6207	gtcctgggggcgcgcgcgc	4257	4276	15
SEQ ID NO: 5662	ccagggtatgggtctctatc	1738	1757	SEQ ID NO: 6208	gataacttccctacgcgc	5084	5103	15
SEQ ID NO: 5663	gcacacgcgcgcgcgcgc	1894	1913	SEQ ID NO: 6209	ggcgccgcgcgcgcgcgc	6343	6362	15
SEQ ID NO: 5664	caagcagtcgcgcgcgcgc	2026	2045	SEQ ID NO: 6210	gggtgtctccctcctgtg	6919	6938	15
SEQ ID NO: 5665	acgtgcgcgcgcgcgcgc	2030	2049	SEQ ID NO: 6211	acgggcgcgcgcgcgcgc	4202	4221	15
SEQ ID NO: 5666	ccgggaagcaccgcgcgcgc	2101	2120	SEQ ID NO: 6212	ggcgctgtatgcgcgcgc	3886	3905	15
SEQ ID NO: 5667	agccacgcgcgcgcgcgc	2115	2134	SEQ ID NO: 6213	catatgtcccaatggcct	3137	3156	15
SEQ ID NO: 5668	gtatgtgggggggcgcgcgc	2235	2254	SEQ ID NO: 6214	tcaagtgtgccatctaca	4011	4030	15
SEQ ID NO: 5669	gagtgagcgcgcgcgcgc	2354	2373	SEQ ID NO: 6215	agggcgaggggtggcgactc	3400	3419	15
SEQ ID NO: 5670	ctcctgtgcgcgcgcgcgc	2474	2493	SEQ ID NO: 6216	ccaccctatgtgcgcgcgc	8861	8880	15
SEQ ID NO: 5671	agccacgcgcgcgcgcgc	2566	2585	SEQ ID NO: 6217	ggcgctccacagtcgaagct	7834	7853	15
SEQ ID NO: 5672	ggcgccatgtgtctctat	2698	2717	SEQ ID NO: 6218	atagaagaagcgcgcgcgc	7865	7884	15
SEQ ID NO: 5673	ggcgcatatgtctctat	2699	2718	SEQ ID NO: 6219	catagaagaagcgcgcgcgc	7864	7883	15
SEQ ID NO: 5674	ggcatatgtctctatgtg	2700	2719	SEQ ID NO: 6220	ccatagaagaagcgcgcgc	7863	7882	15
SEQ ID NO: 5675	gtcgtgtgtgggttcccc	2913	2932	SEQ ID NO: 6221	ggggggcgcgcgcgcgcgc	6402	6421	15
SEQ ID NO: 5676	cccccctcgcgcgcgcgcgc	2926	2945	SEQ ID NO: 6222	cccccctatgtgaacgggc	9376	9395	15
SEQ ID NO: 5677	ggcgagcgcgcgcgcgcgc	3401	3420	SEQ ID NO: 6223	ggagggcgcgcgcgcgcgc	8066	8085	15
SEQ ID NO: 5678	atgtgtgactgtctaccat	3574	3593	SEQ ID NO: 6224	atgtgtacgcgcgcgcgcgc	4157	4176	15
SEQ ID NO: 5679	gtgtgtgactgtctaccat	3575	3594	SEQ ID NO: 6225	catgttgcgcgcgcgcgcgc	4156	4175	15
SEQ ID NO: 5680	gtgtgtgtgtgtgtgtgtgt	3695	3714	SEQ ID NO: 6226	gcacagtcctgtgtgtgt	8543	8562	15
SEQ ID NO: 5681	acacatgtgtgtgtgtgtgt	3704	3723	SEQ ID NO: 6227	ggcgcggtgtgtgtgtgtgt	6342	6361	15
SEQ ID NO: 5682	ccacatgtgtgtgtgtgtgt	3705	3724	SEQ ID NO: 6228	ctgcgcgcgcgcgcgcgcgc	6341	6360	15
SEQ ID NO: 5683	ggcatgtgtgtgtgtgtgt	4325	4344	SEQ ID NO: 6229	ccaggtgtgtgtgtgtgtgt	4979	4998	15

SEQ ID NO: 5684	bagcgagagcggctggagc	4347	4366	SEQ ID NO: 6230	gctccccagcgctgctt	5804	5823	15
SEQ ID NO: 5685	ggagcgcggtgttctgtgc	4361	4380	SEQ ID NO: 6231	gcaaggagcagcgccctcc	7443	7462	15
SEQ ID NO: 5686	cgaaagccatcaagggggga	4489	4508	SEQ ID NO: 6232	loccccagcgctgctgtc	5806	5825	15
SEQ ID NO: 5687	tggaagtgctctatcagcgc	5165	5184	SEQ ID NO: 6233	gcccaggtatcaatctctca	7225	7244	15
SEQ ID NO: 5688	gggtgctgttagagcgagtg	5322	5341	SEQ ID NO: 6234	actcgcgatccaccacc	8765	8784	15
SEQ ID NO: 5689	gtgggttaggtactctgtt	5390	5409	SEQ ID NO: 6235	acaaacalgttctacgccac	7713	7732	15
SEQ ID NO: 5690	cgccgagcaattcaagcag	5515	5534	SEQ ID NO: 6236	ctgcagcgtctcccgcg	6550	6569	15
SEQ ID NO: 5691	tggaagtcgaagtggcgagc	5592	5611	SEQ ID NO: 6237	gctctctatcaggaattca	8175	8194	15
SEQ ID NO: 5692	tgggagcgctttggagacct	5603	5622	SEQ ID NO: 6238	agggtgcctgatcacgcca	7633	7652	15
SEQ ID NO: 5693	gcccgtctccaccacagaa	5739	5758	SEQ ID NO: 6239	ttctggcggtgtatggggc	5895	5914	15
SEQ ID NO: 5694	tgagtgactctcaagacctg	6306	6325	SEQ ID NO: 6240	caggctataaaatgctca	8363	8382	15
SEQ ID NO: 5695	atgtcaaaaagcgttccat	6456	6475	SEQ ID NO: 6241	atgttaccgacccatacat	4158	4177	15
SEQ ID NO: 5696	cgaaaaaactctcagcaaca	6488	6507	SEQ ID NO: 6242	tggtcttccaatgtgtcgg	8708	8727	15
SEQ ID NO: 5697	ggcgccgctcaagctcaag	6565	6584	SEQ ID NO: 6243	cttgaagcgtctcgccgc	8505	8524	15
SEQ ID NO: 5698	gcccctcttgaggcgacaca	6967	6986	SEQ ID NO: 6244	gtctctctatgtgaaggcg	3814	3833	15
SEQ ID NO: 5699	caccgcgtggagctggag	7078	7097	SEQ ID NO: 6245	ctccggtggtacacgggtg	9278	9297	15
SEQ ID NO: 5700	ggaggggggatggaatgaa	7138	7157	SEQ ID NO: 6246	ltcatgtgtgctactcc	7326	7345	15
SEQ ID NO: 5701	gcgcgatcatccatgtgg	7202	7221	SEQ ID NO: 6247	ccaggggggggagggcgcg	9150	9169	15
SEQ ID NO: 5702	tgccaaactctcaaggccc	7301	7320	SEQ ID NO: 6248	ggggcgccacttgggcga	9162	9181	15
SEQ ID NO: 5703	ccccctctgagggggagc	7520	7539	SEQ ID NO: 6249	gctcccgcgctagtgggg	645	664	15
SEQ ID NO: 5704	ctgtctgctaactgtctact	7606	7625	SEQ ID NO: 6250	gttaggactctgcaaggcgag	4809	4828	15
SEQ ID NO: 5705	catggcagcaggtgacctgat	7626	7645	SEQ ID NO: 6251	atcatgtgaacgactccatg	8996	9015	15
SEQ ID NO: 5706	atggagacggtgacctgatc	7627	7646	SEQ ID NO: 6252	galcatgtgaacgactccat	8995	9014	15
SEQ ID NO: 5707	gctatgatcagtagtaactc	8635	8654	SEQ ID NO: 6253	ggagcaactgaaanaagcg	8920	8939	15
SEQ ID NO: 5708	ccactatgactcaactccct	27	46	SEQ ID NO: 6254	agggcgcttgcaacatggg	785	804	24
SEQ ID NO: 5709	agctgtctacactctctgcc	1206	1225	SEQ ID NO: 6255	ggcggtctgacgactagct	8459	8478	24
SEQ ID NO: 5710	ctgcaatgactccctccag	1624	1643	SEQ ID NO: 6256	ctgtgctgggtctgtggag	5847	5866	24
SEQ ID NO: 5711	atgtggggggcgctgagca	2238	2257	SEQ ID NO: 6257	gtctgcgccatcaacaat	7701	7720	24
SEQ ID NO: 5712	ggggacatctctctgggc	3322	3341	SEQ ID NO: 6258	gcccacactgtctccccca	5795	5814	24
SEQ ID NO: 5713	gggacatcatctctggcct	3324	3343	SEQ ID NO: 6259	aggcaggagataacttccc	5076	5095	24
SEQ ID NO: 5714	gggagatactctctgggccc	3366	3385	SEQ ID NO: 6260	ggcgagcgtcagccttccc	6545	6564	24
SEQ ID NO: 5715	atgtgtgactcttccatcat	3574	3593	SEQ ID NO: 6261	atgtctctacgcacagacat	7718	7737	24
SEQ ID NO: 5716	ccagccttaccatcaccca	6189	6208	SEQ ID NO: 6262	tggttacaaggagtgctcg	6382	6401	24
SEQ ID NO: 5717	gcccctctgaggggcgaca	6967	6986	SEQ ID NO: 6263	tgccccagggggggggggc	9147	9166	24
SEQ ID NO: 5718	ccagccccgatggggggc	1	20	SEQ ID NO: 6264	gcgcagggcgagggctcg	550	569	14
SEQ ID NO: 5719	accatagatcactccctctg	28	47	SEQ ID NO: 6265	cagggccttggcacatgtg	784	803	14
SEQ ID NO: 5720	atgagtgctgtgcagcttc	95	114	SEQ ID NO: 6266	gagcgcgagtgccatcat	2948	2966	14
SEQ ID NO: 5721	gtgcagcctccaggagcccc	104	123	SEQ ID NO: 6267	gggggagcgcatcatgcac	6403	6422	14
SEQ ID NO: 5722	tgagcctccaggaccccc	105	124	SEQ ID NO: 6268	ggggggagcgcatcatgca	6402	6421	14
SEQ ID NO: 5723	ccaggacccccccctccgg	113	132	SEQ ID NO: 6269	ccggctgtgtctgtgtcg	9255	9274	14
SEQ ID NO: 5724	accocccctccgggagag	118	137	SEQ ID NO: 6270	ctctcatgccaacgtgggt	6368	6387	14
SEQ ID NO: 5725	ccccctcccgaggagccca	121	140	SEQ ID NO: 6271	tgccaatgagggcagtgagg	598	617	14
SEQ ID NO: 5726	agactctgacgcagtagt	243	262	SEQ ID NO: 6272	actatgogggtcccggtct	3953	3972	14
SEQ ID NO: 5727	agccgagtagtggtggctc	251	270	SEQ ID NO: 6273	gaccagactctgtctggct	3656	3675	14
SEQ ID NO: 5728	gtgtctgtcgagctgcccg	299	318	SEQ ID NO: 6274	cgggcgctgtgttgtaacc	2139	2158	14
SEQ ID NO: 5729	gcgagtgccccgggaggtc	306	325	SEQ ID NO: 6275	gaccocggcgtaggtgcg	671	690	14
SEQ ID NO: 5730	accgtgcacatgagacag	331	350	SEQ ID NO: 6276	cggtcaatctgttacggt	2437	2456	14
SEQ ID NO: 5731	ccccggcggtgtgtatgatc	412	431	SEQ ID NO: 6277	gatcatgcatctccggg	997	1016	14
SEQ ID NO: 5732	gcgcgcagggggccccagg	451	470	SEQ ID NO: 6278	ctctgcacgccttccccgc	6549	6568	14
SEQ ID NO: 5733	accocctgggaaggcgacag	511	530	SEQ ID NO: 6279	ctgtatgaccocgggggggt	3891	3910	14

SEQ ID NO: 5734	ccgcgggaagcgacagc	512	531	SEQ ID NO: 6280	gctgtatgaccocggggg	3890	3909	14
SEQ ID NO: 5735	agcctatcccaaggctcg	528	547	SEQ ID NO: 6281	cgaggcgaggcgctgggt	553	572	14
SEQ ID NO: 5736	atccccaaggctcgccg	531	550	SEQ ID NO: 6282	cggtgtgtgtcccgatg	5418	5437	14
SEQ ID NO: 5737	atccccaaggctcgccg	532	551	SEQ ID NO: 6283	ccggctgtgtgtccgata	5417	5436	14
SEQ ID NO: 5738	cggtgtatcttggccctc	577	596	SEQ ID NO: 6284	gaggcgccagcgagccg	8067	8086	14
SEQ ID NO: 5739	gcatggggtggcgagatg	609	628	SEQ ID NO: 6285	catcgatgacctccatgc	706	725	14
SEQ ID NO: 5740	ctgtaccoccgcggtc	630	649	SEQ ID NO: 6286	gagctgcgaagctccagga	8523	8542	14
SEQ ID NO: 5741	ggggcccccaggaaccccg	661	680	SEQ ID NO: 6287	ccggcgccatgtcgggccc	4064	4083	14
SEQ ID NO: 5742	ggccccaaggaaccccg	662	681	SEQ ID NO: 6288	gcccgcgcgatgtcgggcc	4063	4082	14
SEQ ID NO: 5743	cggtctgcgcgacctcatg	724	743	SEQ ID NO: 6289	catgaggtatcatcgggcgc	6472	6491	14
SEQ ID NO: 5744	ggcctgcgcgacctcatg	725	744	SEQ ID NO: 6290	ccatgaggtatcatcgggcc	6471	6490	14
SEQ ID NO: 5745	ggcccccagggggcgctg	764	783	SEQ ID NO: 6291	cagctccgaattgtcgcc	7414	7433	14
SEQ ID NO: 5746	ggcaccatgtgttcgggt	792	811	SEQ ID NO: 6292	accacgctgcacgggcca	5188	5207	14
SEQ ID NO: 5747	cttctctgttgctctgtg	868	887	SEQ ID NO: 6293	cagcataggcttgggaag	5863	5882	14
SEQ ID NO: 5748	catgtcacgaacagctct	944	963	SEQ ID NO: 6294	agcagctgtacctctcatg	6847	6866	14
SEQ ID NO: 5749	gaggcgccgggaactgatca	983	1002	SEQ ID NO: 6295	gatggcatcatcagccctc	5712	5731	14
SEQ ID NO: 5750	catcccccctacgacaata	1096	1115	SEQ ID NO: 6296	tattaccggggtctgatg	4592	4611	14
SEQ ID NO: 5751	gctgttccactctcgccc	1207	1226	SEQ ID NO: 6297	gggctgtgtgggaacacgc	8793	8812	14
SEQ ID NO: 5752	ggcccgccggcgcgcgaca	1223	1241	SEQ ID NO: 6298	gtctctactatgaaggcg	3814	3833	14
SEQ ID NO: 5753	ggcctggggaactgatgat	1293	1312	SEQ ID NO: 6299	atcaattgtcttccctgcca	5981	6000	14
SEQ ID NO: 5754	cacaagccgtcatcgacat	1362	1381	SEQ ID NO: 6300	atgtttgggactgggtgtg	6279	6298	14
SEQ ID NO: 5755	agcgtcatgcagactgtg	1366	1385	SEQ ID NO: 6301	ccaccaagcaggcgagggt	5560	5579	14
SEQ ID NO: 5756	ggtggcgccggcccgactgg	1381	1400	SEQ ID NO: 6302	ccaggctgcaggcgagcc	5127	5146	14
SEQ ID NO: 5757	ggggcgccactggggagtc	1387	1406	SEQ ID NO: 6303	gactagtgactccggccccc	8641	8660	14
SEQ ID NO: 5758	atggcggggaactgggcta	1430	1449	SEQ ID NO: 6304	tagcgtgtctcacttccat	6846	6865	14
SEQ ID NO: 5759	ttgtgtgtgtgtgtacttt	1454	1473	SEQ ID NO: 6305	aaagcaagcgtgccatcaa	7665	7684	14
SEQ ID NO: 5760	caacgggggggcagcgtgc	1500	1519	SEQ ID NO: 6306	gcagaagcgctctgggttg	5530	5549	14
SEQ ID NO: 5761	acgtctgccgcctcaccag	1512	1531	SEQ ID NO: 6307	ctggaccggaggagagcgt	2278	2297	14
SEQ ID NO: 5762	cagagaatccagcttata	1564	1583	SEQ ID NO: 6308	tatatcgggggtactccgta	8393	8412	14
SEQ ID NO: 5763	ccaatggcattgttgcaca	1586	1605	SEQ ID NO: 6309	gtgtcctcggggcotttgt	2132	2151	14
SEQ ID NO: 5764	ccaatggcagttggccat	1587	1606	SEQ ID NO: 6310	atgttgctcggggccttgg	2131	2150	14
SEQ ID NO: 5765	gtctatcacttatgtcta	1749	1768	SEQ ID NO: 6311	tcaggactggggtaaggac	4176	4195	14
SEQ ID NO: 5766	ctgagcctacaaaagcccc	1764	1783	SEQ ID NO: 6312	gggtgtgctcatgcctcag	9063	9082	14
SEQ ID NO: 5767	caggltgtgtgtccagtgt	1844	1863	SEQ ID NO: 6313	acactccagtttaactctg	8817	8836	14
SEQ ID NO: 5768	gtgtgtccagtgattgtct	1850	1869	SEQ ID NO: 6314	agcaggggccalcaaccaca	7949	7968	14
SEQ ID NO: 5769	gcttaccocccagctcgt	1866	1885	SEQ ID NO: 6315	acagcagaggcggttaagc	6887	6906	14
SEQ ID NO: 5770	ctgtgtgtgtggggacaac	1881	1900	SEQ ID NO: 6316	gttgcaacttggcagacag	2295	2314	14
SEQ ID NO: 5771	gccgcgcgaaggcaactgg	1972	1991	SEQ ID NO: 6317	ccagttgacttcatccgc	9241	9260	14
SEQ ID NO: 5772	gcgaactgtgtgtgtgta	1982	2001	SEQ ID NO: 6318	tacacgggtgcccattgoc	7287	7306	14
SEQ ID NO: 5773	gcgaactgtgtgtgtgta	1983	2002	SEQ ID NO: 6319	gtacacgggtgcccattgoc	7286	7305	14
SEQ ID NO: 5774	ccccgtgttaacatcggggg	2043	2062	SEQ ID NO: 6320	ccccaatcgtatgaacgggg	9376	9395	14
SEQ ID NO: 5775	ggactgtctcgggaagcac	2092	2111	SEQ ID NO: 6321	gtgtcgtgtaggcggagctc	5324	5343	14
SEQ ID NO: 5776	gagctgtcgtgggaagcacc	2093	2112	SEQ ID NO: 6322	gtgtcgtgtaggcggagctc	5323	5342	14
SEQ ID NO: 5777	ccgggaagcaccocggagcc	2100	2119	SEQ ID NO: 6323	gctactcaggtcttcacgga	8616	8635	14
SEQ ID NO: 5778	actcaaaatgtgtctcggg	2124	2143	SEQ ID NO: 6324	ccggcgccagcggtgtgagt	8201	8220	14
SEQ ID NO: 5779	ggctgtgtgtgcaccctag	2142	2161	SEQ ID NO: 6325	ctagcgcggcccaaaaggcc	3611	3630	14
SEQ ID NO: 5780	aggagagcgttgtcaactgt	2287	2306	SEQ ID NO: 6326	caagcgtgtatgggtctct	8162	8181	14
SEQ ID NO: 5781	ggacagctcggagactcagc	2314	2333	SEQ ID NO: 6327	gctgggggtctatgttgc	3128	3147	14
SEQ ID NO: 5782	cagatcggactcagctcagccg	2317	2336	SEQ ID NO: 6328	gggtgtggccactgtctgt	3837	3856	14
SEQ ID NO: 5783	ggagctcagcccgctgtgt	2323	2342	SEQ ID NO: 6329	cagctgtgtgaagagctcc	6206	6225	14

SEQ ID NO: 5784	ccacctacccgctctgcc	2383	2402	SEQ ID NO: 6330	ggactgggtgtgcacgggtg	6286	6305	14
SEQ ID NO: 5785	cggtctgtctcaclggctt	2391	2410	SEQ ID NO: 6331	aagcaggcgaggaggtgcgcg	5564	5583	14
SEQ ID NO: 5786	cccatcagaacatctgtggac	2419	2438	SEQ ID NO: 6332	gtccctgttagtgcaltgg	3929	3948	14
SEQ ID NO: 5787	ggtagcagcgggtgtctctt	2460	2479	SEQ ID NO: 6333	aaggtatgtatgtatgacc	8875	8894	14
SEQ ID NO: 5788	ggcgcccttagagaacactgg	2579	2598	SEQ ID NO: 6334	ccagttgacacttctccgc	9241	9260	14
SEQ ID NO: 5789	gccttagagaaactcgggtg	2582	2601	SEQ ID NO: 6335	ccaccaagcaggcgaggc	5559	5578	14
SEQ ID NO: 5790	ggcgagagcgcacggcatcc	2621	2640	SEQ ID NO: 6336	ggatgtggcgccacgcgcgc	3214	3233	14
SEQ ID NO: 5791	gctgtcatctgtcgggagcg	2786	2805	SEQ ID NO: 6337	ggccacgacatccgcgcgc	7726	7745	14
SEQ ID NO: 5792	attattgacactgtgcca	2824	2843	SEQ ID NO: 6338	ggcacaacagacgtctaat	4647	4666	14
SEQ ID NO: 5793	tcgcataatatacaagtggt	2837	2856	SEQ ID NO: 6339	acacaatcttctcggca	3539	3558	14
SEQ ID NO: 5794	cgccattattacaagtglt	2838	2857	SEQ ID NO: 6340	aaacacaacttctctggcg	3538	3557	14
SEQ ID NO: 5795	gtcgggggagggccgcgagt	2939	2958	SEQ ID NO: 6341	catcgccacagctcgtggac	4327	4346	14
SEQ ID NO: 5796	ctaccccactcggggattg	3201	3220	SEQ ID NO: 6342	caatttacaatgtgtga	8325	8344	14
SEQ ID NO: 5797	ttgtggccgcgcgcgccta	3217	3236	SEQ ID NO: 6343	aggctcaggggcgccgtccaa	5221	5240	14
SEQ ID NO: 5798	ctacgggacactcgggtag	3233	3252	SEQ ID NO: 6344	ctactctacttctctgag	9338	9357	14
SEQ ID NO: 5799	ccgtgtgtctctctcaga	3260	3279	SEQ ID NO: 6345	tgctctacacatggacagg	7617	7636	14
SEQ ID NO: 5800	ctgtgtcttctctgacat	3261	3280	SEQ ID NO: 6346	atgtctctacacatggacag	7616	7635	14
SEQ ID NO: 5801	ccctggggggcgagcacccgc	3297	3316	SEQ ID NO: 6347	ggcgggtaggagctggcagg	4804	4823	14
SEQ ID NO: 5802	ggggggcgagcacccgcggcg	3301	3320	SEQ ID NO: 6348	cgcccaactcgtctccccc	5794	5813	14
SEQ ID NO: 5803	ggcggtgtgtggagacatcat	3316	3335	SEQ ID NO: 6349	gatgttctcgggtgcgc	3755	3774	14
SEQ ID NO: 5804	tgggcgccgcgcgcgatagct	3378	3397	SEQ ID NO: 6350	agacgacgcacgcgtgccca	4761	4780	14
SEQ ID NO: 5805	ggaccacagtcggaggggag	3499	3518	SEQ ID NO: 6351	ctcacctatgtggcaagttc	4222	4241	14
SEQ ID NO: 5806	ggagggggagggttcaagtg	3509	3528	SEQ ID NO: 6352	ccacctgtcaagcccccct	7304	7323	14
SEQ ID NO: 5807	aggcccaactcggccagatg	3625	3644	SEQ ID NO: 6353	catccgcagcggcgggcct	7734	7753	14
SEQ ID NO: 5808	ggcccaatctgcacatagtg	3626	3645	SEQ ID NO: 6354	acatcccgacgcgcggcgc	7733	7752	14
SEQ ID NO: 5809	caggatctctcgtcgctggc	3659	3678	SEQ ID NO: 6355	ggcaataggccatttcctg	9410	9429	14
SEQ ID NO: 5810	aggatctcgtcggctggcc	3680	3699	SEQ ID NO: 6356	ggccaataggccatttcct	9409	9428	14
SEQ ID NO: 5811	ggcccccgcggcgcggttcc	3682	3701	SEQ ID NO: 6357	ggaacctcagcagcagggc	7938	7957	14
SEQ ID NO: 5812	gcacctgtggcagctcggga	3711	3730	SEQ ID NO: 6358	ccggtgtgtacacgggtgc	7279	7298	14
SEQ ID NO: 5813	ctgtggcagctcggaccit	3715	3734	SEQ ID NO: 6359	aaggcacaagggtctcacag	7826	7845	14
SEQ ID NO: 5814	gcggggcgacataagagg	3775	3794	SEQ ID NO: 6360	ccctgcctgtggaccccgc	5682	5701	14
SEQ ID NO: 5815	ggaggtgtctctcccccag	3792	3811	SEQ ID NO: 6361	ctggttgggtcacagctcc	6806	6825	14
SEQ ID NO: 5816	ggaggtctgtctccccagg	3793	3812	SEQ ID NO: 6362	ctggttgggtctcacagct	6805	6824	14
SEQ ID NO: 5817	acttgaaaggcctctcggg	3822	3841	SEQ ID NO: 6363	ccggtgtgttgatcccaagt	5585	5604	14
SEQ ID NO: 5818	gtccccgttgatgcatg	3928	3947	SEQ ID NO: 6364	catgtctacgccaagaca	7717	7736	14
SEQ ID NO: 5819	gaaactactatcggtgtccc	3947	3966	SEQ ID NO: 6365	gggaaggcacactatttct	4504	4523	14
SEQ ID NO: 5820	aaactactatcgcggtcccc	3948	3967	SEQ ID NO: 6366	ggggggcgcatatacagttt	4828	4847	14
SEQ ID NO: 5821	ctccactcgtgcagcggcaa	4032	4051	SEQ ID NO: 6367	tgccaggaccactctggag	4993	5012	14
SEQ ID NO: 5822	ggcgatatgtctcaaaagca	4138	4157	SEQ ID NO: 6368	tgctgcaccgcgtacgcc	4377	4396	14
SEQ ID NO: 5823	gcgtatatgtctaaagcac	4139	4158	SEQ ID NO: 6369	gtgctcgcaccgcgtacgc	4376	4395	14
SEQ ID NO: 5824	tgggglaaggacaattacc	4183	4202	SEQ ID NO: 6370	ggtaaccatgtctcccca	6119	6138	14
SEQ ID NO: 5825	accattaccacggcgcgccc	4193	4212	SEQ ID NO: 6371	ggcgctgtgatctgtggt	5833	5852	14
SEQ ID NO: 5826	cgtaactccactatggcaa	4218	4237	SEQ ID NO: 6372	tgcccaaccagatagc	8669	8688	14
SEQ ID NO: 5827	cagtcctggagcaacggga	4335	4354	SEQ ID NO: 6373	tcgttgagccgcgatcagtg	9560	9579	14
SEQ ID NO: 5828	aggggggaagcgcccatcat	4500	4519	SEQ ID NO: 6374	atgagcggcgaggcgccct	5948	5967	14
SEQ ID NO: 5829	catccacgaagaagtgtgcg	4526	4545	SEQ ID NO: 6375	gcgatgactcgagagagtg	9569	9588	14
SEQ ID NO: 5830	atacatgtctgtagcgtatt	4577	4596	SEQ ID NO: 6376	aatagacactggagttgat	8882	8701	14
SEQ ID NO: 5831	catacagcagccagcgagac	4618	4637	SEQ ID NO: 6377	gtctccccacgcacatag	6128	6147	14
SEQ ID NO: 5832	aggactgctcagggcgagg	4811	4830	SEQ ID NO: 6378	ccctgcacatctctctct	5992	6011	14
SEQ ID NO: 5833	gggaacggccctcgggcatt	4857	4876	SEQ ID NO: 6379	atgtctaccgacccctccc	6863	6882	14

SEQ ID NO: 5834	ggggcagctgttcattctc	4869	4888	SEQ ID NO: 6380	gagggccgcaagccagcccg	8067	8086	14
SEQ ID NO: 5835	tggtacgagctcaaccccg	4922	4941	SEQ ID NO: 6381	cggggactgtcccaacca	8662	8681	14
SEQ ID NO: 5836	gggtctactaaataacacc	4962	4981	SEQ ID NO: 6382	gggtgctcatctctagcc	9518	9537	14
SEQ ID NO: 5837	ggttactcaataacacaa	4963	4982	SEQ ID NO: 6383	gggtgctcatctctagcc	9517	9536	14
SEQ ID NO: 5838	gagataacttccctacct	5082	5101	SEQ ID NO: 6384	aggttggccaggggggtc	6908	6927	14
SEQ ID NO: 5839	ccactctccatcggtggat	5140	5159	SEQ ID NO: 6385	atccaaagtttggctatgg	7906	7925	14
SEQ ID NO: 5840	catggcgatcgatgtcgcc	5278	5297	SEQ ID NO: 6386	ggcctctctcagatcatg	9596	9615	14
SEQ ID NO: 5841	ggccgacctgggaagtcgtc	5293	5312	SEQ ID NO: 6387	gagcccccacattcgcc	7885	7904	14
SEQ ID NO: 5842	ggccgacctggaagtcgtca	5294	5313	SEQ ID NO: 6388	tgagcccccacattcgcc	7884	7903	14
SEQ ID NO: 5843	gggaagtcgtcacccagcac	5301	5320	SEQ ID NO: 6389	gtgccatgtcaggttcca	6676	6695	14
SEQ ID NO: 5844	gcacctgggtgtctaggg	5316	5335	SEQ ID NO: 6390	ctctacacatggacaggtgc	7620	7639	14
SEQ ID NO: 5845	gttatctgtgggtagatc	5383	5402	SEQ ID NO: 6391	galcatcggggccgaaacc	6478	6497	14
SEQ ID NO: 5846	cccgatagggaagtcctct	5429	5448	SEQ ID NO: 6392	agagcggctttatctcggg	8383	8402	14
SEQ ID NO: 5847	gaaatggcaaatgcgcc	5461	5480	SEQ ID NO: 6393	ggcgcgctctggccctcca	6924	5943	14
SEQ ID NO: 5848	ccaaagtggcgagctttgga	5598	5617	SEQ ID NO: 6394	lccattgttagagctctgg	7240	7259	14
SEQ ID NO: 5849	ttcatcagcgggatacagt	5645	5664	SEQ ID NO: 6395	actgcacagtgctcgtgaa	8541	8560	14
SEQ ID NO: 5850	agcgggcttataccacctg	5668	5687	SEQ ID NO: 6396	caggggggtgctggcgctg	5913	5932	14
SEQ ID NO: 5851	ccagcccgctcaccaccaca	5736	5755	SEQ ID NO: 6397	ggcgcgctglatcgtgctg	5832	5851	14
SEQ ID NO: 5852	gtggcgctgtgtatcgtcg	5831	5850	SEQ ID NO: 6398	cagcaggggccatcaaccac	7948	7967	14
SEQ ID NO: 5853	ggaaagtgctagtgcacat	5877	5896	SEQ ID NO: 6399	atgtggtctcacaccttcc	8152	8161	14
SEQ ID NO: 5854	ggatcagtgagcggcgagcgc	5944	5963	SEQ ID NO: 6400	cgccctctcagaccagacc	7453	7472	14
SEQ ID NO: 5855	catgtggcccgcggaagagg	6056	6075	SEQ ID NO: 6401	ccctcttgaggggcgacatg	6965	6984	14
SEQ ID NO: 5856	atgtggcgccggggagaggg	6057	6076	SEQ ID NO: 6402	ccctctctgaggggcgacat	6968	6987	14
SEQ ID NO: 5857	ggggcgctgcagtgatga	6074	6093	SEQ ID NO: 6403	catgctctctctatcgcc	7505	7524	14
SEQ ID NO: 5858	gggtctgctgtcgcggtga	6104	6123	SEQ ID NO: 6404	taccaccagagctatccgc	2751	2770	14
SEQ ID NO: 5859	ggggtaaacatgtctccoc	6117	6136	SEQ ID NO: 6405	gggggagccgggggacccc	7531	7550	14
SEQ ID NO: 5860	catcacccagctgctgaag	6199	6218	SEQ ID NO: 6406	cttcagagcggagggggatg	7130	7149	14
SEQ ID NO: 5861	aggagctgtccagccgtg	6240	6259	SEQ ID NO: 6407	cacggcgacccgcccctcc	7444	7463	14
SEQ ID NO: 5862	tcaagacctggctccagt	6314	6333	SEQ ID NO: 6408	actgcacagtgctcgtgaa	8541	8560	14
SEQ ID NO: 5863	ctctctgcgcggttaccgg	6338	6357	SEQ ID NO: 6409	ccggggagctgtcaaggag	7804	7823	14
SEQ ID NO: 5864	caccacggccgcccctgcacg	6538	6557	SEQ ID NO: 6410	ctgtggagctgcagcgggtg	6613	6632	14
SEQ ID NO: 5865	ggaggtctacgggggtggg	6616	6635	SEQ ID NO: 6411	ccccctcaattaccacctcc	7317	7336	14
SEQ ID NO: 5866	gaggttcacgggggtgggg	6617	6636	SEQ ID NO: 6412	ccccctctgaccagacctc	7455	7474	14
SEQ ID NO: 5867	atgtcaggtgtccagctctc	6682	6701	SEQ ID NO: 6413	aggagatggcgcggaacat	7059	7078	14
SEQ ID NO: 5868	atgaataatccattgcgcg	7152	7171	SEQ ID NO: 6414	gcccgtgatgggctctcat	8165	8184	14
SEQ ID NO: 5869	ctccattgttagagcttg	7239	7258	SEQ ID NO: 6415	caagtggcgagcttggag	5599	5618	14
SEQ ID NO: 5870	tgccattctgacactgtca	7295	7314	SEQ ID NO: 6416	tgactaatctcaaaagggca	8409	8428	14
SEQ ID NO: 5871	accactccacggagaaaa	7327	7346	SEQ ID NO: 6417	ttttttccctcttattggt	9502	9521	14
SEQ ID NO: 5872	ccaactccacggagaaaa	7328	7347	SEQ ID NO: 6418	ttttccctcttattggtg	9504	9523	14
SEQ ID NO: 5873	actccacggagaaaaagg	7330	7349	SEQ ID NO: 6419	ccitttgacagactgcaggt	7770	7789	14
SEQ ID NO: 5874	gggtgtcctgacggactcc	7351	7370	SEQ ID NO: 6420	ggagctgcgtaccacaaac	7390	7409	14
SEQ ID NO: 5875	ccgtaccagacctccgaca	7460	7479	SEQ ID NO: 6421	gttactacaactgtgacagg	7617	7636	14
SEQ ID NO: 5876	agcaagctgccatcaacg	7667	7686	SEQ ID NO: 6422	cggttgacacacttctgtc	7686	7705	14
SEQ ID NO: 5877	ggatgaccattaccgggac	7792	7811	SEQ ID NO: 6423	gtccagatttgagacttacc	9238	9257	14
SEQ ID NO: 5878	ggcaagaagaatgagtttt	8028	8047	SEQ ID NO: 6424	aaaaagcccttgatgtgcca	8931	8950	14
SEQ ID NO: 5879	ggcaagaagaatgagtttt	8029	8048	SEQ ID NO: 6425	gaaaaagcccttgattgcc	8930	8949	14
SEQ ID NO: 5880	gggcagcgggtctgatttc	8204	8223	SEQ ID NO: 6426	gggaagaagaacagctgcc	7680	7699	14
SEQ ID NO: 5881	gactagctgcggttaatacc	8470	8489	SEQ ID NO: 6427	ggatccgccttgcgagtc	9091	9110	14
SEQ ID NO: 5882	ctcgcgtatccaccacccc	8768	8785	SEQ ID NO: 6428	gggggtacgccttgcgag	9089	9108	14
SEQ ID NO: 5883	agagtgattctgatgacc	8876	8895	SEQ ID NO: 6429	gggttcagcggtgtctctc	2458	2477	14

SEQ ID NO: 5884	agccaacttgacacaccca	8976	8995	SEQ ID NO: 6430	gagataaataagggtgact	9052	9071	14
SEQ ID NO: 5885	gggagccgcccctgcgagt	9090	9109	SEQ ID NO: 6431	actcgcgatccaccaccc	8765	8784	14
SEQ ID NO: 5886	ctgcaatgactccctccag	1624	1643	SEQ ID NO: 6432	ctggcgggcatggggcgag	5897	5916	3
SEQ ID NO: 5887	ccagccccgattgggggc	1	20	SEQ ID NO: 6433	gccacacggggagctcgtg	1391	1410	23
SEQ ID NO: 5888	aaggcgacagcctatcccc	520	538	SEQ ID NO: 6434	gggggctccccctcctt	6918	6937	23
SEQ ID NO: 5889	ggccaccagggaacccccgc	662	681	SEQ ID NO: 6435	ggcccaagctgicaggcc	4553	4572	3
SEQ ID NO: 5890	ggcgccggcgagatgatca	983	1002	SEQ ID NO: 6436	gataacatcatgttctc	8697	8716	23
SEQ ID NO: 5891	ctgcaatgttgatctac	1249	1268	SEQ ID NO: 6437	gtaggcggtctctcgag	5330	5349	23
SEQ ID NO: 5892	ctccagactgggtttctg	1637	1656	SEQ ID NO: 6438	caagtgccgagctttgag	5599	5618	23
SEQ ID NO: 5893	tgctactcgtctgcaggt	1830	1849	SEQ ID NO: 6439	acctcagatcattgaacga	9199	9008	23
SEQ ID NO: 5894	caagactgctggcgccccc	2026	2045	SEQ ID NO: 6440	ggggcgaggccgccactg	8586	8595	23
SEQ ID NO: 5895	aatgctgcatgcaactgga	2264	2283	SEQ ID NO: 6441	lccagccacataggccatt	9405	9424	23
SEQ ID NO: 5896	caacctaccggctctgtcc	2383	2402	SEQ ID NO: 6442	ggactacgtccctccggtg	7267	7286	23
SEQ ID NO: 5897	cgccatattacaagggtgt	2838	2857	SEQ ID NO: 6443	aaagccaccacgaaggcg	5554	5573	23
SEQ ID NO: 5898	cgagagccatacaagggggga	4489	4508	SEQ ID NO: 6444	lccagatattgggagttcg	8097	8116	23
SEQ ID NO: 5899	ccagcccgctcaccaccca	5736	5755	SEQ ID NO: 6445	gggtacaagggaagtctcg	6382	6401	23
SEQ ID NO: 5900	ggctatgactgagtactcc	8635	8654	SEQ ID NO: 6446	ggagacatataccagcc	9284	9303	23
SEQ ID NO: 5901	ctccaccatagatcactcc	24	43	SEQ ID NO: 6447	ggagacatcgggccagag	9111	9130	13
SEQ ID NO: 5902	lccaccatagatcactccc	25	44	SEQ ID NO: 6448	ggggagtctgatgaattga	5451	5470	13
SEQ ID NO: 5903	paccatagatcactccct	27	46	SEQ ID NO: 6449	agggcgcccaaggtgggtg	458	477	13
SEQ ID NO: 5904	lcactcccctgtagggagc	36	55	SEQ ID NO: 6450	gttctgaggagcggcgta	809	828	13
SEQ ID NO: 5905	cggttagtatgagtgctg	88	107	SEQ ID NO: 6451	cacgtctcagcgggccaacg	5191	5210	13
SEQ ID NO: 5906	tgctgtgcagcctccagga	100	119	SEQ ID NO: 6452	ctcgtgtcgtggggaca	1879	1898	13
SEQ ID NO: 5907	ccccccctcccgaggagc	119	138	SEQ ID NO: 6453	gctcccgccgtatgtggg	645	664	13
SEQ ID NO: 5908	ggagagccatagtggtctg	131	150	SEQ ID NO: 6454	cgatcattgaacagcctc	8993	9012	13
SEQ ID NO: 5909	gagccatagtggtctggg	134	153	SEQ ID NO: 6455	ccgctgctgggtgagcctc	1048	1067	13
SEQ ID NO: 5910	gtgtctgcgggaaccggtg	142	161	SEQ ID NO: 6456	caccocatattagatccac	5038	5057	13
SEQ ID NO: 5911	agtcacccggcaattgcag	161	180	SEQ ID NO: 6457	ctgcggcgcttgcctact	1408	1425	13
SEQ ID NO: 5912	ggctcttcttggtatcaac	188	207	SEQ ID NO: 6458	gttagtgactcaagacc	6304	6323	13
SEQ ID NO: 5913	ttcttggtatcaaccgctc	194	213	SEQ ID NO: 6459	gagcgaggagggtgagaa	7134	7153	13
SEQ ID NO: 5914	clcaaatgcgtggagatttg	210	229	SEQ ID NO: 6460	caaagactctcgacgtctg	7486	7505	13
SEQ ID NO: 5915	tgctctggagattgggcgt	215	234	SEQ ID NO: 6461	acgcggccgccgcaaggca	1967	1986	13
SEQ ID NO: 5916	gctctggagattggcgctg	216	235	SEQ ID NO: 6462	cacgcggccgcgcgaaggc	1966	1985	13
SEQ ID NO: 5917	gagatttgaggctgcccc	221	240	SEQ ID NO: 6463	ggggacaaccgatcgctc	1891	1910	13
SEQ ID NO: 5918	aaaggccttggtgactgc	273	292	SEQ ID NO: 6464	gcagaagaaggctacattt	7756	7775	13
SEQ ID NO: 5919	aaggccttggtgactgcg	274	293	SEQ ID NO: 6465	ggcagaagaaggctacatt	7755	7774	13
SEQ ID NO: 5920	gtgtactcgtctgatagg	282	301	SEQ ID NO: 6466	ccclaccggctgtctcac	2385	2404	13
SEQ ID NO: 5921	ctgtatagggtgcttgcga	291	310	SEQ ID NO: 6467	tcgcggcccgaggggcagg	544	563	13
SEQ ID NO: 5922	cgatgctcccgaggagctct	307	326	SEQ ID NO: 6468	agacgcagtgctcgctcg	4780	4799	13
SEQ ID NO: 5923	gtcccgagggtctctgat	312	331	SEQ ID NO: 6469	ctacttaggtttggcgcc	4122	4141	13
SEQ ID NO: 5924	lactctgttgcgcgcagg	442	461	SEQ ID NO: 6470	ctcgtctgctggagggttaa	1023	1042	13
SEQ ID NO: 5925	lactctgttgcgcgcagg	443	462	SEQ ID NO: 6471	ccctcgtctggagggttaa	1022	1041	13
SEQ ID NO: 5926	ctgtgttgcgcgcaggggc	445	464	SEQ ID NO: 6472	gccccgaagccagacagg	8348	8367	13
SEQ ID NO: 5927	ctgtgttgcgcgcaggggcc	446	465	SEQ ID NO: 6473	ggccccgaagccagacag	8347	8366	13
SEQ ID NO: 5928	lccgagcggctgcgaacccc	497	516	SEQ ID NO: 6474	ggggccaaggagcttcggga	7922	7941	13
SEQ ID NO: 5929	ggctgcgaaccccgtggag	504	523	SEQ ID NO: 6475	ctctctgcagatggagacc	3268	3287	13
SEQ ID NO: 5930	gtcgcgaaccccgtggag	505	524	SEQ ID NO: 6476	ctctacacattgagacgac	4749	4768	13
SEQ ID NO: 5931	aaagcgacagcctatcccc	520	539	SEQ ID NO: 6477	ggggcgctgccaggcgctt	774	793	13
SEQ ID NO: 5932	cagcctatccccagggtcc	527	546	SEQ ID NO: 6478	gagcacaggcttaatgctg	2252	2271	13

SEQ ID NO:	5633	gagggcaggggcctgggctc	554	573	SEQ ID NO:	6479	gagcgtcttcaacagcctc	5020	5039	13
SEQ ID NO:	5834	cagggcctgggctcagccc	559	578	SEQ ID NO:	6480	gggacatcgacacagctcgt	4324	4343	13
SEQ ID NO:	5835	ggggcctgggctcagcccg	561	580	SEQ ID NO:	6481	ccggcgcacatgcccggcc	4064	4083	13
SEQ ID NO:	5836	ctcggcctcagccgggta	564	583	SEQ ID NO:	6482	taccgacccataacacagg	4162	4181	13
SEQ ID NO:	5837	ccccctctggcaatgag	590	609	SEQ ID NO:	6483	ccctgcgcgacctcagggg	727	746	13
SEQ ID NO:	5838	gagggcctcaggggtggcag	606	624	SEQ ID NO:	6484	ctggcgactgttttctc	1180	1199	13
SEQ ID NO:	5839	agggcctcaggggtggcag	606	625	SEQ ID NO:	6485	ccgtcgtcttcaacacacct	2370	2389	13
SEQ ID NO:	5840	aggagtgccctcgttcaacc	622	641	SEQ ID NO:	6486	gggtcagcgggtgtctcct	2459	2478	13
SEQ ID NO:	5841	gatggctcgttcaaccccg	624	643	SEQ ID NO:	6487	cggggggcgtctacgacatc	4261	4280	13
SEQ ID NO:	5842	gttcaaccccgggctcccg	633	652	SEQ ID NO:	6488	cggggcgcggttccctgaca	3688	3707	13
SEQ ID NO:	5843	gtcaaccccgggctcccg	634	653	SEQ ID NO:	6489	ccggggcgcggttccctgac	3687	3706	13
SEQ ID NO:	5844	gcggctccggcctagttg	642	661	SEQ ID NO:	6490	caacgctccggggaggcgcc	2935	2954	13
SEQ ID NO:	5845	ctcccgccctagttggggc	646	665	SEQ ID NO:	6491	gcctcttgacacactggag	4439	4458	13
SEQ ID NO:	5846	ataccctcacatcgccct	711	730	SEQ ID NO:	6492	aggcaacatatacatgtat	8839	8858	13
SEQ ID NO:	5847	ttccgctcgcgcggcgccc	750	769	SEQ ID NO:	6493	gggcaaaagcacatctggaa	6625	5644	13
SEQ ID NO:	5848	ccctcagggggcgctgcc	767	786	SEQ ID NO:	6494	tggaatgagggcctagggg	508	617	13
SEQ ID NO:	5849	gtcaacagggaacctgcc	832	851	SEQ ID NO:	6495	gggctcatctgctcagca	3092	3111	13
SEQ ID NO:	5850	cgctaaacgcctcggggta	922	941	SEQ ID NO:	6496	taccacccagagctctacgc	2751	2770	13
SEQ ID NO:	5851	tcagcactgtgtttgagg	968	987	SEQ ID NO:	6497	ccctctatgcgccctctga	7512	7531	13
SEQ ID NO:	5852	cccaacgctcggcgccagg	1070	1089	SEQ ID NO:	6498	tcgtgttaaacatcttggg	5763	5782	13
SEQ ID NO:	5853	cgcccgaggatgtaccat	1080	1099	SEQ ID NO:	6499	atggcactgcaltgctggcg	5279	5298	13
SEQ ID NO:	5854	acgacataatgcacacacg	1106	1125	SEQ ID NO:	6500	ctgtggggacaaacagatgt	1888	1907	13
SEQ ID NO:	5855	ggggcggtctctctctgc	1140	1159	SEQ ID NO:	6501	gagcaacttgaaaaagccc	8921	8940	13
SEQ ID NO:	5856	ctgtgggggaacctctcgga	1168	1187	SEQ ID NO:	6502	tcgtgtcgggagcgccag	2615	2634	13
SEQ ID NO:	5857	agctgttcaacctctcgc	1206	1225	SEQ ID NO:	6503	ggcgacataatgagggagct	3719	3798	13
SEQ ID NO:	5858	ctgttcaacctctcgc	1208	1227	SEQ ID NO:	6504	ggggagacataatacag	9282	9301	13
SEQ ID NO:	5859	ctgcaattgttgatctac	1249	1268	SEQ ID NO:	6505	glaggactcgccaggggcag	4809	4828	13
SEQ ID NO:	5860	attgttgatctaccccg	1254	1273	SEQ ID NO:	6506	ccggcccaaaaaggcccaat	3615	3634	13
SEQ ID NO:	5861	atctaccccgccacacgct	1262	1281	SEQ ID NO:	6507	acgcgatggagccggagat	2766	2785	13
SEQ ID NO:	5862	cgccacacgcctcaggtcac	1270	1289	SEQ ID NO:	6508	gtgactctattttgcg	1460	1479	13
SEQ ID NO:	5863	ccgcatggcctgggacatg	1288	1307	SEQ ID NO:	6509	catggaaactactatcgcg	3943	3962	13
SEQ ID NO:	5864	cgcaattctccggaatcc	1344	1363	SEQ ID NO:	6510	gggaaccaccaggagatg	8693	8612	13
SEQ ID NO:	5865	cccaacagccgctcatcgac	1360	1379	SEQ ID NO:	6511	gtcgtcaccagcacctggg	5306	5325	13
SEQ ID NO:	5866	ctggggagctcgtcgccggc	1396	1415	SEQ ID NO:	6512	ccggcgagcgcatggccag	1695	1714	13
SEQ ID NO:	5867	ggcgggccgttgcctactat	1408	1427	SEQ ID NO:	6513	atagaagaagccgtccgccc	7865	7884	13
SEQ ID NO:	5868	tttgcgggctgtgacgggc	1472	1491	SEQ ID NO:	6514	ggccccaacttggccaaa	7888	7907	13
SEQ ID NO:	5869	caacctcaacaggggggg	1492	1511	SEQ ID NO:	6515	ccccataatgcaggagggt	4420	4439	13
SEQ ID NO:	5870	ggggggggcagcgtccccc	1504	1523	SEQ ID NO:	6516	gcggcagcgcgacggccccc	7440	7459	13
SEQ ID NO:	5871	ggggcgacgctccgcctc	1507	1526	SEQ ID NO:	6517	gagggagctgtcctccccc	3789	3808	13
SEQ ID NO:	5872	gcccgcctcaacagcggtt	1517	1536	SEQ ID NO:	6518	accctcaaacggggggggc	1493	1512	13
SEQ ID NO:	5873	atccagctataaaacacca	1571	1590	SEQ ID NO:	6519	gggtatcgtgtggtaggat	5382	5401	13
SEQ ID NO:	5874	ctccagactgggtttcttg	1637	1656	SEQ ID NO:	6520	caacggcgagacggctggag	4346	4365	13
SEQ ID NO:	5875	ccgggagcgcatggccagc	1696	1715	SEQ ID NO:	6521	gtctgtggcgctctccggc	3869	3888	13
SEQ ID NO:	5876	ctcggcctccatgacaaag	1714	1733	SEQ ID NO:	6522	cttgttaccatacaaggagc	2667	2686	13
SEQ ID NO:	5877	aagttgcaccaggatggg	1730	1749	SEQ ID NO:	6523	cccaaccagaatacagactt	8673	8692	13
SEQ ID NO:	5878	gggggtctatactatgac	1746	1765	SEQ ID NO:	6524	gcgtgtgtgtgttcccc	2914	2933	13
SEQ ID NO:	5879	ccagaggctcttatctdgg	1786	1805	SEQ ID NO:	6525	ccagagctctcgtctcgtg	3658	3677	13
SEQ ID NO:	5880	cccaacctcaacatgttgt	1810	1829	SEQ ID NO:	6526	accaagatcatcaccttgg	3284	3303	13
SEQ ID NO:	5881	tcgtacgtcgtcgcaggt	1830	1849	SEQ ID NO:	6527	accttcaacatgagacga	4748	4767	13
SEQ ID NO:	5882	tgctcgtcaggtgtgttgt	1837	1856	SEQ ID NO:	6528	acatgtctccccacgca	6123	6142	13

SEQ ID NO:	5983	gggggcaaacagatgctct	1890	1909	SEQ ID NO:	6529	agacagcagccgtgcccca	4761	4780	13
SEQ ID NO:	5984	cagctgggggggaagacgat	1924	1943	SEQ ID NO:	6530	atcgagcgctcagcccgctg	2320	2339	13
SEQ ID NO:	5985	gcgcgcgaagcgaactgggt	1974	1993	SEQ ID NO:	6531	aaccagggaggtatcgcgcg	8596	8615	13
SEQ ID NO:	5986	gcgcgaagggcaactggttc	1975	1994	SEQ ID NO:	6532	gaacccagggaggtatcgcg	8595	8614	13
SEQ ID NO:	5987	ctgtacatggatgaatagc	1996	2015	SEQ ID NO:	6533	gctataaaatcgctcacag	8386	8385	13
SEQ ID NO:	5988	gtacatggatgaatagca	1997	2016	SEQ ID NO:	6534	tgctgctcaatgtcttoca	7607	7626	13
SEQ ID NO:	5989	gttccaccaagactgtcggg	2020	2039	SEQ ID NO:	6535	cccgctcacaccaccagaac	5740	5759	13
SEQ ID NO:	5990	agacgtgcccggggccccc	2028	2047	SEQ ID NO:	6536	ggggaggtgtcaagtggctt	3512	3531	13
SEQ ID NO:	5991	cccccggtgaacatgcgggg	2042	2061	SEQ ID NO:	6537	ccccaatgatgaagcggg	9376	9395	13
SEQ ID NO:	5992	taacacctgacctgcgcc	2071	2090	SEQ ID NO:	6538	ggggagcgaactgttgctta	8561	8580	13
SEQ ID NO:	5993	ggctctggcactacccctg	2184	2203	SEQ ID NO:	6539	caggaggtatcgcgcgagcc	8600	8619	13
SEQ ID NO:	5994	gcactgtgcaactcttoca	2201	2220	SEQ ID NO:	6540	tggtatgggggtcggttgca	6717	6736	13
SEQ ID NO:	5995	cagggttaattgctgcatgc	2257	2276	SEQ ID NO:	6541	gcatcatgacacaccactg	6411	6430	13
SEQ ID NO:	5996	aattgctgcatgcaadgga	2264	2283	SEQ ID NO:	6542	ccatggttgctagcgatt	9009	9028	13
SEQ ID NO:	5997	ctgcatgcaadctggacccg	2268	2287	SEQ ID NO:	6543	cgggacactgtgatgcag	3236	3255	13
SEQ ID NO:	5998	caactgaccccgaggagag	2275	2294	SEQ ID NO:	6544	ctctatcgggatgaggttg	6761	6780	13
SEQ ID NO:	5999	gacagggacagatcgagac	2309	2328	SEQ ID NO:	6545	gctctccccagggcctgct	3799	3818	13
SEQ ID NO:	6000	gacagatcggtgactcagcc	2315	2334	SEQ ID NO:	6546	ggctggagcgcggtgtgtc	4357	4376	13
SEQ ID NO:	6001	acagatcggtgactcagccc	2316	2335	SEQ ID NO:	6547	ggggcacaagccctgctgt	5201	5220	13
SEQ ID NO:	6002	actggctgtatccaccctcc	2402	2421	SEQ ID NO:	6548	ggagaggggcggtgcagt	6068	6087	13
SEQ ID NO:	6003	ggctgtgtatccactccatc	2405	2424	SEQ ID NO:	6549	gatgatgctgctgalagcc	2551	2570	13
SEQ ID NO:	6004	gtcagcggtgtgtctctctt	2461	2480	SEQ ID NO:	6550	aaaggacggtgtgctcagc	7344	7363	13
SEQ ID NO:	6005	gagtagtgtgtgtgtgtt	2492	2511	SEQ ID NO:	6551	aaagacaagctcacaactc	9202	9221	13
SEQ ID NO:	6006	tlggatgatgctgctgat	2547	2566	SEQ ID NO:	6552	ctcactgtggcattcaca	5707	5726	13
SEQ ID NO:	6007	ccgagcgccgcttagagaa	2574	2593	SEQ ID NO:	6553	tlctgattgctcactatcgg	3015	3034	13
SEQ ID NO:	6008	agaaacactgtgggcctcaa	2589	2608	SEQ ID NO:	6554	ttgatatacacaactctc	3000	3019	13
SEQ ID NO:	6009	tacatcaagggcaggctgg	2672	2691	SEQ ID NO:	6555	ccagatgtacacdaatgta	3637	3656	13
SEQ ID NO:	6010	caaggcgaggctgtgtcct	2677	2696	SEQ ID NO:	6556	aggggttaggcactcactg	9355	9374	13
SEQ ID NO:	6011	gcatgtgcogdgtctctgc	2720	2739	SEQ ID NO:	6557	gcaatgctcacttccatcg	6848	6867	13
SEQ ID NO:	6012	calggccgctgtctctgct	2721	2740	SEQ ID NO:	6558	agcagtgctcacttccatg	6847	6866	13
SEQ ID NO:	6013	ggcctgtctctgtctctc	2725	2744	SEQ ID NO:	6559	gagggcgccactgtgcgc	9160	9179	13
SEQ ID NO:	6014	ggagatggctgcatogtgc	2779	2798	SEQ ID NO:	6560	gcacggcgagcccccctcc	7443	7462	13
SEQ ID NO:	6015	atggctgcatgtgctggag	2783	2802	SEQ ID NO:	6561	ctcaggcccaataggccat	9404	9423	13
SEQ ID NO:	6016	ggcgccggttttgtgggtc	2801	2820	SEQ ID NO:	6562	gaccattaccagcgccgc	4192	4211	13
SEQ ID NO:	6017	ttctatcacacagctgtgag	2887	2906	SEQ ID NO:	6563	ctcacaggccgggacaaga	3482	3501	13
SEQ ID NO:	6018	gtgtgtgggttccccccctca	2918	2937	SEQ ID NO:	6564	gaggttcaactcacaacac	5242	5261	13
SEQ ID NO:	6019	ccccccctcaacgtccgg	2926	2945	SEQ ID NO:	6565	ccggctctgtggctgaggga	6261	6280	13
SEQ ID NO:	6020	ctcaacgtgcggggaggcc	2933	2952	SEQ ID NO:	6566	ggcctgttactcactgatg	8959	8978	13
SEQ ID NO:	6021	acaaacdtgtatgtcca	3008	3027	SEQ ID NO:	6567	ggctctctcagatgttgt	8130	8149	13
SEQ ID NO:	6022	caaacctctgattgccata	3010	3029	SEQ ID NO:	6568	atgatcagcccgcttgttt	8267	8286	13
SEQ ID NO:	6023	ggaccgctcactgtgtctcc	3032	3051	SEQ ID NO:	6569	ggagatcctgcgggaagtc	7171	7190	13
SEQ ID NO:	6024	gacgcctcatgtgtgtcca	3033	3052	SEQ ID NO:	6570	tggaatactatctatgcgtc	3945	3964	13
SEQ ID NO:	6025	atgcatgtatgtgcggaaa	3106	3125	SEQ ID NO:	6571	tttctgtatgggttagcat	9348	9367	13
SEQ ID NO:	6026	tatgttccaaattggccttc	3139	3158	SEQ ID NO:	6572	gaagccagacagagctataa	8354	8373	13
SEQ ID NO:	6027	ccaaattgacctcatgaga	3145	3164	SEQ ID NO:	6573	ttctcagcagagggctctgg	7552	7571	13
SEQ ID NO:	6028	cccttcatgagactggcgcc	3153	3172	SEQ ID NO:	6574	gcgctctgtgctctcaagg	5927	5946	13
SEQ ID NO:	6029	ccctgcggttagcagtgagg	3241	3260	SEQ ID NO:	6575	ctccgcccggaagggaagg	3349	3368	13
SEQ ID NO:	6030	tgtgtctctctctgacatg	3262	3281	SEQ ID NO:	6576	catgtgtctaacgcacagca	7717	7736	13
SEQ ID NO:	6031	tgggggggcagacaccgcgg	3299	3318	SEQ ID NO:	6577	ccgcttcatgtattctcca	8083	8102	13
SEQ ID NO:	6032	ggggggggcagacaccgcgg	3300	3319	SEQ ID NO:	6578	ggcgccaactcgctctccc	5792	5811	13

SEQ ID NO: 6033	gggggacatcatcctggg	3321	3340	SEQ ID NO: 6579	cccatctacacgtctccac	4020	4039	13
SEQ ID NO: 6034	gggggacatcatctgggc	3322	3341	SEQ ID NO: 6580	cccatctacacgtctccac	4019	4038	13
SEQ ID NO: 6035	gggggacatcatctgggc	3323	3342	SEQ ID NO: 6581	ggcccgaggggtctccccc	6913	6932	13
SEQ ID NO: 6036	acdtgtctccgcccgaag	3343	3362	SEQ ID NO: 6582	cccttgacagacgtcagg	7770	7789	13
SEQ ID NO: 6037	gtctccgcccgaaggga	3346	3365	SEQ ID NO: 6583	ccccggtcttcacagaca	3962	3981	13
SEQ ID NO: 6038	ggggagatactctggggcc	3366	3385	SEQ ID NO: 6584	ggcccatctacacgtctcc	4018	4037	13
SEQ ID NO: 6039	ctcccaacagacccggggc	3439	3458	SEQ ID NO: 6585	gcccccccttgagggggag	7519	7538	13
SEQ ID NO: 6040	ctccaccgcacaacattt	3530	3549	SEQ ID NO: 6586	aagaggtctccacacgtlga	6215	6234	13
SEQ ID NO: 6041	caacaattcttctggcgac	3540	3559	SEQ ID NO: 6587	gtctccgagctctgtgtg	6020	6039	13
SEQ ID NO: 6042	ggctggccgggcccgcgc	3671	3690	SEQ ID NO: 6588	cggtgtgtgcaaacacgc	5542	5561	13
SEQ ID NO: 6043	ccccggggcggtctccgt	3685	3704	SEQ ID NO: 6589	cagggttgtaactccgggg	4840	4859	13
SEQ ID NO: 6044	ctcttgacacacgtcacct	3698	3717	SEQ ID NO: 6590	aggctacgcgggtggggga	6618	6637	13
SEQ ID NO: 6045	ctccgggtgcgcggggggg	3762	3781	SEQ ID NO: 6591	ccccgttgatctcatggaa	3931	3950	13
SEQ ID NO: 6046	ctccccagggcctgtctcc	3802	3821	SEQ ID NO: 6592	ggagacgtctccggcaggag	9111	9130	13
SEQ ID NO: 6047	gggggttgcaaaaggcggtg	3904	3923	SEQ ID NO: 6593	caacctgctggggaaacccc	5680	5699	13
SEQ ID NO: 6048	ttgttcccggttgagttcaa	3926	3945	SEQ ID NO: 6594	gggagacctctggggcaaa	5613	5632	13
SEQ ID NO: 6049	cogtaacggcgccatttcca	3996	4015	SEQ ID NO: 6595	ggagtttgtaactccgggg	8940	8959	13
SEQ ID NO: 6050	caagtggtcccatctcaacg	4013	4032	SEQ ID NO: 6596	cggtggtagtagtactcttg	5389	5408	13
SEQ ID NO: 6051	caagctcccatggcagcg	4028	4047	SEQ ID NO: 6597	cgctgtctgggtctctgtg	5815	5834	13
SEQ ID NO: 6052	cgcgtatgtggcgcccaagg	4068	4087	SEQ ID NO: 6598	cccttaaggtctagagcgg	5937	5956	13
SEQ ID NO: 6053	cgatattgtctaaagcaca	4140	4159	SEQ ID NO: 6599	gtggagagtgctctatacg	5163	5182	13
SEQ ID NO: 6054	gtatattgtctaaagcaca	4141	4160	SEQ ID NO: 6600	atgtggaagtgctctatac	5151	5170	13
SEQ ID NO: 6055	ggacacattacacggggcgc	4191	4210	SEQ ID NO: 6601	gcgcgtgtcatctcaggctc	6167	6186	13
SEQ ID NO: 6056	ccccattatcgtctctcac	4209	4228	SEQ ID NO: 6602	gtgggcgcggggagaggggg	8059	8078	13
SEQ ID NO: 6057	agctctctgcgcagcggtg	4236	4255	SEQ ID NO: 6603	ccacagctcaaggctaaact	7839	7858	13
SEQ ID NO: 6058	gagacggctctggcgccggc	4352	4371	SEQ ID NO: 6604	ggccgggggaccccgatctc	7537	7556	13
SEQ ID NO: 6059	cacgtctacgctcctcagga	4384	4403	SEQ ID NO: 6605	ctctacacatggacaggtg	7619	7638	13
SEQ ID NO: 6060	ggagagatgtctacgtctac	4453	4472	SEQ ID NO: 6606	gtacgagctgctcacttcca	6845	6864	13
SEQ ID NO: 6061	agccatcccatcgaagacc	4477	4496	SEQ ID NO: 6607	ggctggttgctgtctggt	9257	9276	13
SEQ ID NO: 6062	ctcccatcgaagccatcaa	4482	4501	SEQ ID NO: 6608	ttgagggggagccggggga	7527	7546	13
SEQ ID NO: 6063	ccccatcgaagccatcaag	4483	4502	SEQ ID NO: 6609	cttgagggggagccggggg	7528	7547	13
SEQ ID NO: 6064	ggcctcggaattcaatgtgt	4568	4587	SEQ ID NO: 6610	cagctccgaattgtccggcc	7414	7433	13
SEQ ID NO: 6065	gtccgtcataccgacacgc	4612	4631	SEQ ID NO: 6611	gctggggatgtttgggac	6271	6290	13
SEQ ID NO: 6066	gtcataccgacccagcggag	4616	4635	SEQ ID NO: 6612	ctccattgagccactgtac	8968	8987	13
SEQ ID NO: 6067	cggtgctataccggtgactt	4668	4687	SEQ ID NO: 6613	aagtcgaagaggtctcccg	7184	7203	13
SEQ ID NO: 6068	ctttgactcagtgatcgac	4684	4703	SEQ ID NO: 6614	gtcgggttctcgtgttaaag	8213	8232	13
SEQ ID NO: 6069	acagtcgactcagctgtg	4724	4743	SEQ ID NO: 6615	ccaaatctacgggggctgt	8947	8966	13
SEQ ID NO: 6070	cttggtcccccacactcaac	4738	4757	SEQ ID NO: 6616	gggtgttgagctctcaag	6301	6320	13
SEQ ID NO: 6071	gagacagcagcgtgcccc	4760	4779	SEQ ID NO: 6617	ggggacaacacgatgtctc	1891	1910	13
SEQ ID NO: 6072	ggggtaggactggcagggg	4806	4825	SEQ ID NO: 6618	ccccccggggacttgcgcc	8657	8676	13
SEQ ID NO: 6073	gggcatacaggtgttgta	4831	4850	SEQ ID NO: 6619	tacaaatggacaggttgccc	7622	7641	13
SEQ ID NO: 6074	gggggacgcggccctcgcc	4855	4874	SEQ ID NO: 6620	gccccctgacgctctcccc	6546	6565	13
SEQ ID NO: 6075	ggacggggctgtgtcttg	4906	4925	SEQ ID NO: 6621	ccaattgacacccacgtcca	8008	8027	13
SEQ ID NO: 6076	ggacggggctgtgtcttg	4907	4926	SEQ ID NO: 6622	accaattgacacccacgtc	8008	8027	13
SEQ ID NO: 6077	gttggtgtgacgagctcac	4918	4937	SEQ ID NO: 6623	gggtgcgctgtgtggcagca	5849	5868	13
SEQ ID NO: 6078	ggccactctctgtccacg	5050	5069	SEQ ID NO: 6624	ctgggcgcgctgacgggca	3164	3183	13
SEQ ID NO: 6079	gggtggcataccaagccaca	5101	5120	SEQ ID NO: 6625	gtgacacaaattgacacc	8002	8021	13
SEQ ID NO: 6080	ggggtcagggcccaacctcc	5130	5149	SEQ ID NO: 6626	ggagggcgcaagccggccgc	8066	8085	13
SEQ ID NO: 6081	ccatcgtgtggacaaatgt	5147	5166	SEQ ID NO: 6627	acattctgtgcgggctatgg	5892	5911	13
SEQ ID NO: 6082	ctatcaggtctaaacccacc	5175	5194	SEQ ID NO: 6628	gtggcctcaaatgctatga	5933	5952	13

SEQ ID NO:	6083	gctgtataggtaggggc	5214	5233	SEQ ID NO:	6629	gcccgaaacggagtagca	6832	6851	13
SEQ ID NO:	6084	ccaaatacatcatggcatg	5268	5287	SEQ ID NO:	6630	catgccclcaggaaactgg	9072	9091	13
SEQ ID NO:	6085	ggagctctcgcagctcgg	5336	5355	SEQ ID NO:	6631	ccagctgtctgcgcctcc	6955	6974	13
SEQ ID NO:	6086	gcctgacacacaggcgagtg	5364	5383	SEQ ID NO:	6632	acactccaggccaataggc	9401	9420	13
SEQ ID NO:	6087	agcccaacagcaggcgagg	5557	5576	SEQ ID NO:	6633	ctccagtttaactctggct	8820	8839	13
SEQ ID NO:	6088	catgtggaattcaatcagc	5635	5654	SEQ ID NO:	6634	gctgcgccatcacaaatg	7702	7721	13
SEQ ID NO:	6089	ctctatcacacagccgcctc	5728	5747	SEQ ID NO:	6635	gagccgcatgactgcagag	9565	9584	13
SEQ ID NO:	6090	cccgagacacccctcgtgt	5751	5770	SEQ ID NO:	6636	aacactcttgggggggtggg	5771	5790	13
SEQ ID NO:	6091	ctcctgtttaacatcttgg	5762	5781	SEQ ID NO:	6637	ccaatcgatgaacggggag	9378	9397	13
SEQ ID NO:	6092	ctgggggggtgggtagcgg	5777	5796	SEQ ID NO:	6638	cgggcgcaaatctatccaa	6564	6583	13
SEQ ID NO:	6093	gcttcggcttctcgtgggc	5818	5837	SEQ ID NO:	6639	gcccgaaacggagctagca	6832	6851	13
SEQ ID NO:	6094	ctgtggcgctgctglatcgc	5829	5848	SEQ ID NO:	6640	ggagcggcgctgctgacga	8453	8472	13
SEQ ID NO:	6095	cgctgctgctgctgtggc	5845	5864	SEQ ID NO:	6641	gccacgacatcccgagcg	7727	7746	13
SEQ ID NO:	6096	cggtctgttgcgcagcatag	5853	5872	SEQ ID NO:	6642	ctctagactctttagcggc	7111	7130	13
SEQ ID NO:	6097	ggggcgagggtggctggcg	5908	5928	SEQ ID NO:	6643	ggcccaactcgtctccccc	5794	5813	13
SEQ ID NO:	6098	ctggcgctgctgtggcctt	5922	5941	SEQ ID NO:	6644	aaggcgaggcgcaagccag	8063	8082	13
SEQ ID NO:	6099	ggcgcgctgctggcccttc	5923	5942	SEQ ID NO:	6645	gaagggaaggcgaagcga	8062	8081	13
SEQ ID NO:	6100	gagcggcgaggcgccctct	5950	5969	SEQ ID NO:	6646	agagcgtcgtctgctgctc	7596	7615	13
SEQ ID NO:	6101	ggggccggggagagggggc	6060	6079	SEQ ID NO:	6647	gccacatctcacgtcccca	4019	4038	13
SEQ ID NO:	6102	cggtctgataggcttgcctt	6095	6114	SEQ ID NO:	6648	aagcagcgaggagctccg	5564	5583	13
SEQ ID NO:	6103	gtgctgagagcgagcgccg	6146	6165	SEQ ID NO:	6649	ggcgcgcgacagcggcac	7428	7447	13
SEQ ID NO:	6104	atgagagactgttctacgcc	6237	6256	SEQ ID NO:	6650	ggcgggcgagcggcatcat	6399	6418	13
SEQ ID NO:	6105	gtccaaagctcctcgcggcg	6331	6350	SEQ ID NO:	6651	ccgctcctgttggaggagc	7969	7988	13
SEQ ID NO:	6106	acagatgcgcgcagatcgtc	6442	6461	SEQ ID NO:	6652	gacatatactacagcctgt	9287	9306	13
SEQ ID NO:	6107	acgtgtgcagtgaacattcc	6506	6525	SEQ ID NO:	6653	gggaagaaccccgactacgt	7257	7276	13
SEQ ID NO:	6108	ggggccctgcacgccttcc	6544	6563	SEQ ID NO:	6654	gggaagaagcgaegtccgc	7660	7679	13
SEQ ID NO:	6109	agtgcccatgctcaggttcc	6675	6694	SEQ ID NO:	6655	ggaaacagctagacacact	8803	8822	13
SEQ ID NO:	6110	tgcccatgtcaggttccag	6677	6696	SEQ ID NO:	6656	ctggcgcgctgcaggggca	3164	3183	13
SEQ ID NO:	6111	agctcctcagtttttccac	6693	6712	SEQ ID NO:	6657	gtgagagcgtctgctcgt	7593	7612	13
SEQ ID NO:	6112	tcaaggagggtggalgggggt	6708	6727	SEQ ID NO:	6658	accctctctcaagcgttga	8153	8172	13
SEQ ID NO:	6113	pacggagggtggatgggggtg	6709	6728	SEQ ID NO:	6659	cacccctctctcaagcgtg	8152	8171	13
SEQ ID NO:	6114	gacccctccacattacag	6872	6891	SEQ ID NO:	6660	ctgttctctcaacggctc	8278	8297	13
SEQ ID NO:	6115	tggcagggggtctcccc	6911	6930	SEQ ID NO:	6661	gggttgggtagcgcccaa	5782	5801	13
SEQ ID NO:	6116	cttgaggcgacatgcac	6972	6991	SEQ ID NO:	6662	gtgcttaaggagatgaagg	7811	7830	13
SEQ ID NO:	6117	ggagatggcggaacatc	7060	7079	SEQ ID NO:	6663	gatgacccattcttctcc	8887	8906	13
SEQ ID NO:	6118	gagatggcggaacatcaca	7061	7080	SEQ ID NO:	6664	gatgacccattcttctcc	8886	8905	13
SEQ ID NO:	6119	ctagactcttgcagcgc	7112	7131	SEQ ID NO:	6665	ggcggtgctgtagcactag	8457	8476	13
SEQ ID NO:	6120	tacactctctgcagcgt	7113	7132	SEQ ID NO:	6666	agcgaggggtctgttgtcta	7556	7575	13
SEQ ID NO:	6121	agaatgaatataccatgctc	7149	7168	SEQ ID NO:	6667	gcaaaagatgaggttttct	8030	8049	13
SEQ ID NO:	6122	tggcggggagatcctgcg	7164	7183	SEQ ID NO:	6668	gcacagatgcatctggcaa	8730	8749	13
SEQ ID NO:	6123	agcggaggaggctgtggaga	7580	7599	SEQ ID NO:	6669	ctcgtgcgcgaccccgct	9309	9328	13
SEQ ID NO:	6124	gagagcgtctgtctgctc	7594	7613	SEQ ID NO:	6670	gcagtaagaagcaagctca	9197	9216	13
SEQ ID NO:	6125	gtcgtctgtctcaatgt	7601	7620	SEQ ID NO:	6671	acatgtctcagcagacac	7716	7735	13
SEQ ID NO:	6126	tgcccatcacacaalgtt	7704	7723	SEQ ID NO:	6672	accatgtctccccacgca	6123	6142	13
SEQ ID NO:	6127	cagaagaaggtcaccitttg	7757	7776	SEQ ID NO:	6673	caaaagaatgaggttttct	8031	8050	13
SEQ ID NO:	6128	ctctgtgatgaccattcgg	7789	7808	SEQ ID NO:	6674	cgggaacctatcacgagg	7936	7955	13
SEQ ID NO:	6129	ggagctgtcttaaggagatg	7807	7826	SEQ ID NO:	6675	catcgggcaggagcgctcc	9116	9135	13
SEQ ID NO:	6130	aaagaatgaggtttctgc	8032	8051	SEQ ID NO:	6676	gcagaagaaggtcaccittt	7758	7777	13
SEQ ID NO:	6131	gttctgtgtatgcagaag	8110	8129	SEQ ID NO:	6677	cttcatcgtcaggaaact	9069	9088	13
SEQ ID NO:	6132	ggctataaaatcgtccaca	8365	8384	SEQ ID NO:	6678	gtgaaaggtctcgtgagcc	9551	9570	13

SEQ ID NO: 6133	ttctccatccttctagctc	8900	8919	SEQ ID NO: 6679	gagcggagggggatgagaa	7134	7153	13
SEQ ID NO: 6134	gtctcgtgcccgaccccg	9303	9322	SEQ ID NO: 6680	cggggcgcgttcctgaca	3688	3707	13

Table 15. Sequences from human hepatitis C virus (HCV) (Direct Match Type)

	Source	Start Index	End Index		Match	Start Index	End Index	Match #
SEQ ID NO: 5285		9446	9465	SEQ ID NO: 5288		9466	9485	2
SEQ ID NO: 5286		9446	9465	SEQ ID NO: 5289		9465	9484	1
SEQ ID NO: 5287		9447	9466	SEQ ID NO: 5290		9466	9485	1

Table 16. Sequences of Exemplary Gene Targets

5	gi 4502152 ref NM_000384.1 Homo sapiens apolipoprotein B (including Ag(x) antigen) (APOB), mRNA	
	ATTCACCGGACCTGCGGGCTGAGTGCCTTCTCGGTGCTGCCGTGAGGAGCCGCGCCAGCCAGC	
10	CAGGCGCGGAGGCGAGGCGCAGGCGCAGCCAGGAGCCGCCCCACGCGAGCTGGGATGGAGCCGCGC	
	AGGCGCGGCTGCTGGGCGCTGCTGGCGCTGCTGCGCTGCTGCTGCTGGCGGGCGCCAGGCGCG	
15	AAGAGAAATGCTGGAAATGTCAGCCTGCTGCTGCTCAAAGATGCGACCGGATTCAGACACCTCCGGAA	
	GTACACATCAACTATGAGGCTGAGAGTTCCAGTGGAGTCCCTGGAGCTGCTGATTCAGAGGTGCCACC	
20	AGGATCACTGCAAGGTTGAGCTGGAGGTTCCCGCAGCTCTGCACTTCATCCTGAAGACAGCCAGTGCA	
	CCCTGAAGAGGCTGATGCGTTCAACCTGAGGGCAAGCCTTGCTGGAAGAAACCAAGAACTCTGAGGA	
25	GTTCGTGACAGCATCTCCAGGTATGAGCTCAAGCTGGCCATCCAGAGGGAAGCAGGTTTCTCTTAC	
	CCGGAGAAAGATGAACCTACTTACATCTCTGAACATCAAGAGGGGACATATTCTGCGCCTCTGTTCCCC	
30	CAGAGACAGAAAGAACCCAAAGCAAGTGTGTTTCTGGATACGCTGTATGGAAACCTGCTCCACTCATCTTAC	
	CGTCAAGCAGGAAAGGCGCAATGTGGCACAGAAATATCCACTGAAGAGACCTGGGCGAGTGTGATCGC	
35	TTCAAGCCCATCCGACAGGCATCAGCCCACTTCTCTCATCAAGGCATGACCCGCCCTTGTCAACTC	
	TGATCAGCAGCAGCCAGTCTCTCAGTACACACTGGACGCTAAGAGGAAGCATGTGGCAGAAAGCCATCTG	
40	CAGAGGCAACACCTCTTCTGCTCTTCTCCTCAACAAATAGTATGGGATGGTAGCACAGTGACACAG	
	ACTTTGAACTTGAAGACACACAAAGATCAACAGCCGCTTCTTGGTGAAGGACTAAGAAAGTGGGCC	
45	TGCGATTTGAGAGGCCAAATCCACATCACTCCAAAGCAGGCGCAAGCTGTTTGAAGACTCTCAGGA	
	ACTGAAAACTAACCATCTCTGAGCAAAATATCCAGAGAGCTAATCTCTCAATAAGCTGTGTTACTGAG	
50	CTGAGAGGCTCAGTGATGAAGCAGTCACTCTCTTGGCCAGCTGATTGAGGTGTCCAGCCCATCA	
	CTTTACAGCCTTGTTGTCAGTGTGGACAGCTCAGTCTCCACTCATCTCCAGTGGCTGAAACGTGT	
55	GCATGCCAACCCCTCTGATAGATGTGTGTCACTACCTGGTGGCCCTGATCCCGCCAGCCTCAGCACAG	
	CAGCTGCGAGAGATCTTCAACTGGCGAGGGATCAGCGCAGCCGAGCCACTTGTATGCGCTGAGCCACG	
60	CGGTCAACAACTATCATAAAGACAAACCTTACAGGGAGCCAGGAGCTGTGGACATTGTCAATCACTGAT	
	GGAAACAGATTCAAGATGACTGCACTGGGGATGAAGATTACACTATTGTATTGCGGGTCACTGGAAAT	
65	ATGGGCGAAACCATGGACAGTTAATCTCCAGAACTCAAGTCTTCAATCTCAATGTGTCCAAAGTACAA	
	AGGCATCACTGATGATCCAGAAAGCTGCCATCAGGCTCTCGGGAATAGCAAGCTTAAGACACAGGACA	
70	GGAGGTTCTCTTCAGACTTCTTGTGATGATGCTTCTCGGGAGATGAAGGACCTGGCTGCTTATG	
	TGATGAGGAGTCTCTTCAAGCAGATATTAACAAATATGTCAAATCTACCATGGGACAGAACTGAGC	
75	AAGTGAAGAACTTGTGGCTTCCCATATTGCCAATATCTTGAACCTCAGAAGAAATGGATATCCAGATCT	
	GAAAAGTTGATGAAGAAAGCTCTGAAGAAATCTCAACTTCCAACTGTGATGGATCTCAGAAAATCTCT	
80	CGGAACATCAACTCTCAAAATCTGTTCTCTTCCATCACTTGACCCAGCTCAGCCAAATGAAGGGGA	
	ATCTTATATTGATCCAAATAACTACCTTCTTGAAGAAAGCATGTGAAAACCTCAAGCTCTGATTTGG	
85	ATTGTCTTCAGCTGACCTCATCGAGATTGGCTTGGAAAGGAAAGGCTTTGAGCCCAATTTGAAGCTCTT	
	TTTGGGAAGCAGAGGATTTTCCAGACAGTGTCAAACAAGCTTTGTACTGGGTAAATGGTCAAGTCTCTT	
90	ATGGTGTCTTAAGGCTTTAGTGGACCACTTTGGCTATACCAAGATGATAAATCATGAGCAGGATATGGT	
	AAATGGAATATGTCAGTGTGAGAGGCTGATTAAGATTTGAAGATCCAAAGAACTCCCGGAAGCCAGA	
95	GCTTACCTCCGACCTCTGGGAGAGGAGCTTGGTTTGGCCAGTCTCCATGACCTCCAGCTCTCTGGGAAGC	
	TGCTCTGATGGGTGCCCGCATCTGCGAGGGATCCCCAGATGATTGGAGAGGCTCATCAGGAAGGGCTC	
100	AAAGAATGACTTTTCTTCTCACTACATCTTTCATGGAGAAATGCCTTGAACCTCCCATCTGAGCTGATTA	
	CAGTGCAGAAATCTTTCATCTGGAGTCATTGCTCCCGGAGCCAGGCTGGAGTAAATCTGGAAGTAGCCA	
105	ACATGCAAGCTGAACTGCTGTGGCAAAACCTTCGCTGTCTGTGGAGTTTGTGACAAATATGGGCATCATCAT	
	TCCGGACTCTGCTAGGAGTGGGTCCAGATGAACCAACTTCTTCCACAGTGCAGGCTCTGAGAGCTCAT	
110	GTTGCTTAAAGCTGGGAAGCTGAAGTTTATCATCTCTCCCAACAGAGACAGTCAAGCTGCTCAGTGT	
	GAGGCAACATATTACATTTGGTCTCTCAACCAAAACGAGGTGATCCACCTCTCATTTGAGAACAGGCA	

GTCCCTGGTCAGTTTGAAGCAAGCTCTTTCCTGGCCCTGAATTACTGCACCTCAGGCGCTTACTCCAAACGCC
 AGCTCCACAGACTCCGCTCTCTACTATCCGTGACCGGGGACACAGATTAGAGCTGGAACTCAGGCGCTA
 CAGGAGAGATTGAGCAGTATCTGTACAGCGCAACTCATGAGCTCCAGAGAGAGACAGAGCCTTGGTGGA
 TACCCTGAAGTTTGAATCAAGCAGAAGGTGCGAAGCAGACTGAGGCTACCATGACATTCAAATATAAT
 5 CGGAGAGATTGACCTTTGTCAGTGAAGTCCAAATTCGGGATTTTGATGTGACCTCGGAAACATCTCA
 GAGTTAATGATGAATCTACTGAGGGCAAACGCTTACAGACTCACCTTGGACATTCAGAACAGAAATAAT
 TACTGTGCTCCGCTCATGGGCCACTTAAGTTGTGACACAAAGGAAGAAAGAAATACAGGGTGTATT
 TCCATACCCCGTTTGCAGAGCAAGCCAGAAAGTGAATCTCCGCCCACTGGTCCGCTGCCAACTGCTCTC
 TCCAAATGAGCATCATCTGCTACAGCTTATGGCTCCACAGTTTCCAGAGGGTGGCATTTAGATGA
 10 AGAGAGATTGAATTTGAATGGAAACACAGGCACCAATGTAGATACCAAAATATGACATTCCTCAATTTCCCT
 GTGGATCTCTCCGATATCTCTAAGAGCTTGATATGTATGTCTAATAGACTCCTGGATCACAGAGTCCCTG
 AAACAGACATGACTTTCCGGCACGTGGGTTCCAAATTAATAGTTGCAATGAGCTCATGGCTTCAGAAGGC
 ATCTGGGAGTCTCTCTTATACCCAGACTTTGCAAGACCACTCAATAGCTGGAAGGATCTCAACCTCCAG
 AACATGGGATTTGCCAGACTTCCACATCCAGAAACCTCTCTTAAAGCGATGGCGGGTCAAAATATA
 15 CCTTGAACAGAACAGTTTGAATATGAGATTCCTTTGCGCTTTTGGTGGCAATCTCCAGAGATCTAAA
 GATGTTAGAGACTGTTAGGACACAGCCCTCCACTTCAAGCTCTGTGGGATTCATCTGCCATCTCGAGAG
 TTCCAGATCCCTACTTTTACCATCCCAAGTTGTATCAAGTCCAAGTGCCTCTCTGGTGTCTTAGAGAC
 TCTCCAGTAAGTCTACAGCACTTGTACAACTGGTCCGCTCTCTACAGTGGTGGCAAGCCAGACAGACA
 CCAATCTCAGGCTCTCGGCTCGTTACACATGAAGGCTGACTCTGTGGTTGACTGCTTTCTTCAATATGTG
 20 CAAGGATCTGAGAAACAAACATATGACCAAGAATACGTTTCACTATCTGTATGGGTTCTTACGCG
 ACCAATTTCTAGATTGCAATTCMAATTCAGTCTATGTAGAAACCTGGAAACCAACCAAGTCTCAAAAGG
 TTATTAATATTTCAATGCACTCTAGTCTCTGGGACCAAGATGCTGCTCTAGTCTTGGATCTGCAAT
 AACCAAGCAGCATTTGTTTGTCAAAGAGTCAAGATTGATGGGAGTTCAAGATCTCTCTGTTCTATGCTA
 25 AAGGACATATGCGCTGTCTGTGAGAGGGATCTTAACACTGGCGGCTCAATGGAGAGTCCAACTGGAG
 GTTTAACTCTCTTACTCTCCAGGACCAACAGAGTAAAGAGATAGAAAGATGGAACCTCTCTCTCCTC
 ACTCTCACTCTGATCTGCAAGTGGCATCATTAAATACTGCTTCTCCCTAAAGATATGAGAGATCAAGC
 TGACTTTAAATCTGACACCAATGGGAAGTATAAGAACTTTGCCACTCTTCAACAGAGTGAATATGACCTT
 CTCTAAGCAAAATGCACTGCTGCGTTCTGAATATCAGGCTGATTACGAGTCATTGAGGCTCTTCAAGCCTG
 30 CTCTTGTGATCACTAAATTTCCATGGTCTTGTGATTAATGTCTGACATCTTAGGCACTGACAAATTAATA
 GTGGTGCTCACAAGGGCACATGAAGATTGGCAGAGTGAATATCTACAGTGCACAGACCAACTTGAA
 GTGTAGTCTCTGTGCTGGAGAAATGAGCTGAAATGACAGGCTTGGCTCTCTGGGGCATCTATGAATAAT
 ACAACAAATGGCGCTCTCAGGGAACACAAATGCAAAATTCAGTCTGGATGGGAAGCCGCTCACAGAGC
 TATCACTGGGAAGTGCTTATCAGGCAATGATCTTGTGGTGTGACAGCAAAACATTTTCAACTTCAAGGT
 35 CAGTCAAGAGGACTTAAGCTCTCAATGACATGATGGGCTCATATGCTGAAATGAAATTTGACACACACA
 AACACTTGAACTTCCAGGCTTATCAGCTGGACTTCTCTTCAAACTTGACAACTTTACAGCTCTGACACA
 AGTTTATTAAGCAAACTGTAAATTTACAGCTACAGCCCTATTCTCTGCTAATCTATTAAAGCATGACCT
 GAAATACAATGCTCTGGATCTCAACCAATGGGAACCTACGGCTAGAACCCCTGGAAGTGCATGTGGCT
 40 GGTAACTTAAAGGAGCCTACCAAAATATGAAATTAACACACTCATGCTGCTCTCTGCTGCTTAT
 CAGCAAGCTATAAAGCAGACACTGTTGTCAAGGTTCAAGGTTGGGAGTTTGGCTTACGCTTCAACACAGA
 CATCGCTGGGCTGGCTTTCAGCCTATGACATGAGCACAACTATAATTCAGACTCATGCTTTTTCAGCAAT
 GTCTTCCGTTCTGTAAATGGCCCCGTTTACCATGACCATCGATGACATACAAATGGCAATGGGAATCTGC
 CTCTCTGGGGAGACATATCGGCAGCTGTATAGCAAACTCTGCTGTGAAGCAAGCACTCTGTCATTTAC
 45 TTCTCTCATGATATCAAAAGGCTCCCAAGTCATCATCTCTGTGCTAGGAAAGCACTCAGTGCAGCTTT
 GAACACAAGAGTCAGTGCCCTGCTTACTCAGCTGAGCAGAGGCACTGGAACCTCAAGAACCAATTTA
 ACACAATGAATACAGCCAGGACTTGGATGCTTACAACACTAAAGATAAATTTGGCTGTGAGGCTTATCTGG
 AGCAATCTGTGCTGACCTTAACCTTACTAGACTCCCAATTAAGTGCCACTTTTACTCAGTGAAGCCATC
 AATATCATTTGATGCTTTAGAGATGAGAGATGCCGTTGAGAAGCCCCAAGAATTTTACAAATTTGTGCTTTG
 50 TAAAGTATGATAAAGAACCAAGATGTTCACTTCAATTAACCTCCCACTTTTGTAGACTTGTCAAGATTTG
 TGAAGAGGATGCAACCAACTATAGTTGTATGTGGAAGACGTACAGAGAACTGGAAGCATTCAATATTT
 GATCAATTTGTAAAGAAATACAGAGCAGCCCTGGGAAACCTCCACAGCAAGCTTAATGATTACTGTCAAT
 55 CATTTCAATTTGGAGAGACAAGTTTCACTATGCCAGGAGAACTGACTGCTCTCACAAAAGATATAGAAAT
 TACAGAAATGATATACAAATGCAATAGATGATGCCAAATCAACTTTAATGAAAACTCATCTCAACTG
 CAGACATATATGATACAAATTTGATCAGTATATTAAGATAGTTATGATTTACATGATTTGAAATATGCTA
 TTGCTAATATTTATGATGAAATCATTTGAAAAATTAAGAAAGTCTTGATGAGCACTATCATATCTCGTGAA
 TTTAGTAAATCAACTCATGATCTACATTTGTTTATTTGAAATATTTGATTTTAAACAAAGTGAAGTAGT
 ACTGCTATCTGGATTCAAAATGTGGATACTTAGTACCAATCAGAACTCAGATCAAGGAAATCTGACG
 AGCTTTAAGAGACACATACAGAATATAGACATCCAGCACTAGCTGGAAAGTTAAACACACCAATTTAGGCG
 TATTGATGTTAGAGTGCTTTAGATCAATTTGGGAACATAATTTCAATTTGAAAGAAATTAATGATGTTCTT

GAGCATGTCAAACACTTTGTTATAAATCTTATTGGGGATTTTGAAGTAGCTGAGAAAAATCAAAGCCTTCA
 GAGCCAAAGTCCATGAGTTAATCGAGAGGTATGAAGTAGACCAACAATCCAGGTTTAAATGGATAAAT
 AGTAGAGTTGACCCACCAATACAAGTTGAAGGAGACTATTGAGAAGCTAAGCAATGCTTACACCAAGTT
 AAGATAAAAGATTACTTTGAGAAATTTGGTTGGATTATTGATGATGCTGTGAAGAGCTTAATGAAATTTAT
 5 CTTTAAAAACATTCAATTGAAGATGTTAAACAAATTCCTTGACATGTTGATAAAAGAAATTAAGGTCAATTGA
 TTACCACCAAGTTTGTAGATGAAACCAATGACAAATCCGTGAGGTGACTCAGAGACTCAATGGTGAAT
 CAGGCTCTGGAACTACCAAAAAGCTGAAGCAATTAACACTGTTTITAGAGAAACCAAGGCCACAGTTG
 CAGTGTATCTGGAAAGCCTACAGGACACCAAAATTAACCTTAAATCATCAATTGGTTACAGGAGGCTTTAAG
 TTAGGATCTTTGGCTCACATGAAGGCCAAATTCOGAGAGACTCTAGAGATACACAGACCGAATGTAT
 10 CAAATGGACATTGAGCAGGAACCTCAACGATACCTGTCTCTGGTAGGCCAGGTTTATAGACCACTTGTCA
 CCTACATTTCTGATTTGGTGGACTCTGTCTGTCTAAGAACCTTATGACTTTGACAGAGCAATATTCTATCCA
 AGATTGGGCTAAACGATGAAGACATTGGTAGACCAAGGGTTTCACTGTTCTCTGAAATCAAGACCATCCTT
 GGGACAGTCCCTGCTTTGAAGTCAGTCTTCAGGCTCTTCAGAAAGCTACCTTCCAGACACCTGATTTTA
 TAGTCCCTCTACAGCAATTGGAGATTCCATCAGTTCAGATAAACTTCAAAGACTTAAAAAATATAAAAAT
 15 CCCATCCAGGTTTTCACACCCAGAATTACCATTCTTAACACCTTCCACATTCCTTCCTTTACAAATTGAC
 TTTGTGAAATGAAAGTAAAGATCATCAGAACCCATTGACCGAGTGAGAACAGTGAAGTCAAGTGGCCCG
 TTTCCAGATATATATCTCAGGGATCTGAAGGTGGAGGACATTCCTCTAGCGAGAAATCACCTTCCAGGACT
 CCGTTTACCGAARAATCGCAATTCAGAAATTCATATCCCACTCTCAACCTTAATGATTTTCAAGTTCCT
 GACCTTCACATACCAAGATTCCAGCTTCCCAACATCTCACACAAATGAAGTACCTACITTTTGGCAAGC
 20 TATACGATTTCTGAAAAATCCAATCTCCTCTTTTCACTATAGATGCAAAATAGGAAATGGAAATGCAAC
 CACCTCAGCAAAAGCAAGCAGGTATCCAGCTTCCATCACTGCCAAAGGAGAGTCCAAATTAGAAGTTCCTC
 AATTTGATTTTTCAGCAAAATGCACAACTCTCAACCCCTAAGATTAAATCCGCTGGCTCTGAAGGGATGAG
 TGAAGTTTCCGACCAAGTACCTGAGAACGGAGCACTGGGAGTGAATGCTGTTTITGGAAGTGTCTATGA
 25 GGGAAATTCACACACAGTGGCAAGTTTACACACGAAAAAATACACTGGAGGCTTAGTAAATGGAGTAAAT
 GTCAAGATAAACRAATCAGCTTACCTGGATAGCAACACTAAATCTTCCACAAATTTGAACATCCCCCAAC
 TGAATCTCTAGTCAGGCTGACCTGCGCAACGAGATCAAGACACTGTTTGAAGGTGCGCCACATAGCATG
 GACTCTTCTGGAATAAGGCTCATGGAAATGGGCGCTGCCAGACTCTTCAGATGAGGGAACACATGAATCA
 CAAATAGTTTTCACCATAGAAGGACCCCTCACTTCTTTGGAGCTGTCCAAATGAAGTCAATAGCAACACCC
 30 TAAGGATAAACAAAACTTGTTTATGAATCTGGCTCCCTCAACTTTCTAACTTGAATTTCAATATCAAT
 AGTCGATTTCCAGCATGTGGGCCACAGTGTTCTAATCTGCTAAAGGCACTGGCACTGTTTGGAGAAGGGGAG
 GCAAGATTCTACGGAGCATGATGCTCATTTAAATGGAAGGTTATTGGCACTTGTGGCACTTCTCTTT
 TCTTTTTCAGCCAGGCACTTTGAGATCACGGCATCCACAAACAAATGAAGGGAATTTGAAGAAATGCTTTTCC
 ATTAAGGTTTAAACGGAAGATAGACTTCTGAAATAACTATGCACTGTTTCTGAGTCCAGGTCCGAGCA
 35 GCAAGTTGGCAAGTAAATGCTAGGTTCAATCAATTAAGTACAACCAAAATTTCTCTGCTGGAACAAAGC
 AGAACATTATGGAGGCCCATGTAGGAATAAATGGAGAAGCAAAATCTGGATTTCTCTTAAACACTTCTTTTAA
 AATTCCTGAAATGCTTACCTTACCAATAATCACAACTCTCCACTGAAAGATTTCTCTCTATGGGAA
 AAAACAGGCTTGAAGGAATTTCTGAAAAAGCAAGCAATCAATTTGATTTAAGTGTAAAGCTCAGATATA
 AGAAAAACAAACACAGGCATTCCATCAAAATCTTTGGCTGTGCTTTGTGAGTTTATAGTCAAGTCAAGCAT
 40 CAAATCCTTTGACAGGCATTTGAAAAAACAAGAAACAAATGATTAGATTTTGTCAACCAATCTTATAAT
 GAAACAAAAATTAAGTTTGAATAGTACAAGCTGAAAAAATCTCAAGAGAGCTCCCAAGAGCTTTCAAA
 TTTCTGGATACAATGTTCCAGTTGTCAATGTTGAAGTGTCTCCATTCAACATAGATAGATGTGGGCATTCGG
 CTATGTGTTCCAAAAGCAGTCAGCATGCTAGTTTCTCCATCTAGGTTCTGAGCTCGGTGTGCTTCA
 TACACATTAACTCTGCCATCATTAGAGCTGCCAGTCTCTCATGTCCCTAGAATTTCAAGCTTTCTCTTCC
 45 CACATTTTAAGGAATTTGTACCATAGCCATATTTTATCTCTGCGCATGGGCCAATATTCACTATGATTT
 TCTCTTTAAATCAAGTGTATCACTAGTAATACCAATGTGAACTTTTAAACAGTCAGATATTTGTGTTGT
 CATCTCTCTTCTCATCTTCTCATCTGTGATGATGCACTGCAATAGAGGCAACATAGAGTATGCTGCT
 CAGAAAAAGGGGATGAAGTTAGCCACAGCTCTGTCTCTGAGCAACAAATTTGTGGAGGGGTAGTCATAA
 CAGTACTGTGAGCTTACCAACGAAAAATATGGAAGTGTCTGTGCAAAAACCAAGAACCGAAATTCCTCA
 50 ATTTTGAAGTGAATTTCAAGCAAGAACTTAATGGAATACCAAGTCAAAACCTACTGTCTCTTCTTCCCA
 TGGAAATTTAAGTATGATTTCAATTTCTCAATGCTGTACTCTACCGCTAAAGGAGGAGTGAACCAAGCT
 TAGCTTTGGAAGCCTCACTCTTACTTTTCAATGAGTCACTACCAAGAGGAGTGTCAAGGGTTGCGGTT
 CTTTCTCGGGAATTTTCAAGAACTATTGCTAGTGAGGCCAACACTTACTTGAATTTCAAGAGCACACGGT
 CTTCACTAGTGAAGCTGAGGCGCACTTCCAAAATTTGATGATATCTGGAACCTTGAAGTAAAAAGAAAAATTTG
 55 TGGAGAAGCCACACTCAACGATATATTTCCCTCTGGGAGCAGTACGAAAGAACCACTTACAGCTAGAG
 GGCCCTCTTTTTCACCAACGAGAAATACAGCAAGCAAGCCACCTGGAACTCTCTCATGCAAGTGTGAG
 CTCTTGTTCAGGTCATGCAAGTCAGCCAGTCTCTTCCATGATTTTCCCTGACCTTTGGCCAGGAATGTC
 CCTGAATGCTCACTAGAACCCAGAAAGATCAGATGGAATAATGAAGTCCGAGTCTCTCTGCTTCTTCC
 CAGAGCCAGGTGAGCTTTCCAATGACCAAGAAAAAGGCACACCTTGACATTGACAGATCTCTTAGAAGGAC

ACCTAAGGTTCTCTCAAAAATATCATCTACCAGTCTATGACAAGAGCTTATGGGATTTCCTAAAGCTGGA
 TGTAACTACACAGGATCTGGTAGGAGACAGCATCTCTGTTTCAACTGCCTTTGTGTACACCAAAACCCC
 AATGGCTATTCAITTCCTCATCCCTGTAAAAGTTTGGCTGATAAAATTCATTCTCTCGCTGGGCTGAACTTAA
 ATGATCTAAATTCAGTCTTGTCTCATGCCTACGTTCCATGTCCCATTTACAGATCTTCAGGTTCATCTGTG
 5 CAAACTTGTACTTCAGAGAAATACAAATCTATAAGAGCTGAGAACTTCATCAITTCGCTTCAACCTACCA
 ACATCTCCCAGAGGTAAATTCCTCGAAGTTGATGTGTAAACAAAATATTCCTCAACACAGAGACTCTCTGA
 TTCCCTTTTGTAGATAACCGTGCTGAATCTCAGTTAACTGTGCCAGTTCACGCTTCCAAAAGTGT
 TTCAGATTGGCATCTGTGCTTTTGGATCTAAATGTCAGTAGCCAACAGATCGCAGATCTTGAGTTGCCACC
 ATCATCTGGCTCGAGCAGACCATTGAGATTCCCTCCATTAAAGTTCTCCGTGTATAATGCCATTTGGAGCTCG
 10 CTTCCTTTCAAGCACTGACTGCACGCTTTGAGGTAGACTCTCCCGTGTATAATGCCATTTGGAGCTCGAG
 TTTGAAAAACAAAGCAGATTATGTTGAAACAGTCTCGGATTCACATGTCAGCTCAACCGTACAGTTCTTA
 GAAATATGAATCAATGTTTGGGAACACACAAAATCGAAGATGGTACGTTAGCTCTAAGACTAAAGGAA
 CACTTGCACACCGTGACTTCAGTGCAGAAATATGAGAAGATGGCAAATTTGAAGGACTTCAGGAATGGGA
 AGGAAAGCGCACCTCAATATCAAAAGCCAGCGTTCCAGCATCTCCATCTGCGCTACCGAAGAACGAAAG
 15 AAAGGCATCTCCACTCGACGAGCCTCCCCAGCCGTTACCCGATCTCCATCTGCGCTACCGAAGAACGAAAG
 ACTTTTCTAAATGGACCTTCTACTACAGCCCTCAGTCTCTCCAGATAAAAACTCAGCATATTCAAAC
 TGATTTGAGGCTCCGGGAATCTGATGAGGAACTCAGATCAAAGTTAAATGGGAAGAGAGGCACTGTTCT
 GCTTTCTCAACCTCTCTGAAAGACACAGTGCACCAAGGCCACAGGGGTCTTTATGATTAAGTACAGACT
 20 ACCACTGTGGACACACACAGGCTCACCCTGAGAGAAAGTGCTCTCAAGCTGAGAGAAATCTGCAGAACAA
 TGTGTAGTGTGGTTTATCAAGGGCCATTAGGCAAAATGATGATATCGAGCTGGAGTTCCAGAAAGCAGCC
 AGTGCCACCACTGGGACCTACCAAGATGGAAGGACAGGCCAGAAATCTGTACCAAGACTCTTGATCACT
 AGGAAGGCCCAAGCAGTTTCCAGGGACTCAAGGATAACGTTGTGATCGAGTTACTCACTCAAAA
 ATTCATATGAAAGTCAAGCATCTGATTGACTCACTCAITGATTTTCTGAACCTCCCAGATTCAGATTTT
 CCGGGGAACCTGGGATATACACTAGGGAGGAACCTTGCACTATGTTTCAATAGAGGAGTGGAGCGGTAC
 25 TGTCCAGGATATCTGAAAGTCTCAATATGTTGAGAAATACGTTTCTTCTATTTCAAGACTCAGTGAT
 TACACTTCTCTTGAGTTTAAAGAAACATAAATTAATAGATGTAATCTGATGATATAGGAACTGTTTGA
 GATTATTAAGAAAGGCCCAAGAGGATTTTAAAGCCATTCAGTCTCTCAAGACCCACAGAGTGCTACGTA
 ATCTTCAGGACCTTTTACAATTCATTTTCCAATATAGAGATAACATTAACACGCTGAAAGAGATGAA
 30 ATTTACTATTCTTATTAATATATATCCAAGATGAGATCAACAACTCTCAATGATTTATATCCCATGTTGT
 TTTAAATGTTGAAAGAAACCTATGCTCTTAATCTTCATAAGTTCAATGAATTTATTAACAAAGAGCTTC
 AGGAAGCTCTCAAGAGTTACAGCAGATCCATCAATACATTAATGGCCCTCGTGAAGAAATATTTGATCC
 AAGTATAGTTTGGCTGCAGATGAAATATTAATGAATGAAAGAAAGATAGTCAGTCTGATCAAGAACCTG
 TTAGTTGTCTTTCAGGACTCTCCATCTGGAATATATTTGTCAAGTCTCACTTTTACTCTCCCAACTCTCAA
 35 GTCAAGTTGAGCAATTTCTGCACAGAAATATTCAGAAATATCTTAGCATCTTACCGATCCAGATGGA
 AGGGAAGAGAGATTTCAGAGCTTTCTGCCACTCTCAGGAATAATTTAAAGGCCCAAGACTTACGACG
 AAGAAATAATTTCTGATTACCACCAAGCAGTTTAGATATAAATGCAAGATTTTTCAGACCACTCTCTG
 ATTTACTATGAAAAATTTATTTGCTGAATCCAAAAGATTGATTGACCTGTCCATTCAAAACCTACCAAT
 TCTGATATACATCACGGAGTTACTGAAAAGCTGCAATCAACACAGCTCATGAACCCCTACATGAAGCTT
 40 GCTCCAGGAGACTTACTATCATCTCTTAATTTTAAAAAGAAATCTTCATTATTTCTTTTCCAAAT
 GAACTTTTCAGTATGACAGAAAAAATTCACAACTGCTTATGATTAACCACTACAGTAGGACCGCTTG
 CAGTAGGACAGTAGATATAAGCAGAGACACATGAATGAGCTGCACAAAGCTGGCAGACAGGCTTG
 GAGGCTCTGAACTCAGAGGATGGCAATTTTTCAGAGTTAAAGAAATCAGAGCTTGAGTTATTTTGC
 TAAACTTGGGGAGGAGAACAAATAAATGGAGTCTTTATTTGTGTATCATATA (SEQ ID NO: 6681)

45 >gi|4557442|ref|NM_000078.1| Homo sapiens cholesteryl ester transfer protein,
 plasma (CETP), mRNA
 GTGAATCTCTGGGCGCAGGAAGACCTGCTGCTCCCGAGAGCCCTCATGTTCCGTGGGGCTGGGCGGACA
 TACATATACGGCTCTCAGGCTGAAACGCTCGGCGCACTTACACCACTGCCTGATTAACCATGCTGGCTG
 CACAGACTCTCAGACCTGGCCCTGCTGGGCAATGCCATGCCTGCTCAAAGCCACTCGCACGAGCGAGG
 50 CATCGTGGCGCATCACCAAGCTGCTCCTCTGTTGTTGAACACAGAGACTGCAAGGTGATTCAGAGAC
 GCCCTTCCAGCGAGCCAGCTACCCAGATATCAAGGCGAGAGGCCATGATGCTCCTTTGGCCAAAGTCAAGT
 ATGGGTTGCACAACTCTCAGATCAGCCACTGTTCATCGCGCAGCCAGGTGGAGCTGTGTGAAGGCCAA
 GTCCATGTATGTTCTCATTCAGAACGCTGTCTGTGTCTTCAAGGGGACCCGTAAGTCTGCTACACACT
 GCTGTGTGCTGGTGTATGATCAGTCCATGACTTCAGATCGACTCTGCCATGACCTCGAGTCAACAA
 55 CACAGCTGACCTGTGACTCTGTAGAGTGGGACCGATGCCCTGACTGCTACTCTGTTTTCATTAAGCT
 GCTCCTGCTATCTCAAGGGAGCGAGAGCTGGTGTGATCAAGCAGCTGTTCAAAATTTCACTCTCTTC

ACCCTGAAGCTGGTCTCTGAAGGACAGATCTGCAAAAGAGATCAACGTCATCTCTAACATCATGCGCGGATT
 TTGTCGACAGAGGGCTGCCAGCATCTTTTCAGATGGAGACATTGGGGTGGACATTTCCCTGACAGGTGA
 TCCCGTCATCACAGCCTCTTACCTGGAGTCCCATCACAAGGGTCATTTCATCTACAAGAAATGTTCTCAGAG
 5 GAACCTCCCCTCCACCACTTCTCGCCCACTGCTGGGGGACCTCCCGCATGCTGTATCTTCTGGTTCCTCTG
 AGCGAGTCTTCCACTGCTGCTGGCCAAAGGTAGCTTTCCAGGATGGCCGCTCATGCTCCAGCTGATGGGAGA
 CGAGTTCAGGCGAGTCTGAGAGCTTGCGCTTCAACACCAACACAGGAATCTTCCAGAGGTTGTGCGG
 GGCTTCCCAGCCAGGCCCCAAGTCAACGCTCCACTGCTCCAGATGCCAAGATCTCTGCCAAACAAAGG
 GAGTCGTGGTCAATTCTTCAGTGATGGTGAATTCCTCTTTCCACGCCAGACAGCAACATTTCTGTAGC
 10 TTACACATTTGAGAGGATATCGTGACTACCGTCCAGGCTCTCTATTCTAAGAAAGAGCTCTCTTTAAGC
 CTCTCTGGATTTCAGAGTATACACAAAGACTGTGTTCCAACTGACTGAGAGACCTCCAGCTCATTCAGCA
 GCTTCTCTGCASTCAATGATCAACCGCTGTGGGCACTCCCTGAGTCAATGTTCTCGGCTCGAGGTAGTGTTCAC
 AGCCCTCATGAACAGCAAGGGCTGAGGCTCTTTCGACATCATCAACCTGAGATTATCACTCGAGATGGC
 TTCTCTGCTGTCAGATGAGCTTTGGCTTCCCTGAGCACCTGCTGGTGGATTTCCTCCAGAGCTTGAGCT
 AGAAGTCTCAAGGAGGTGCGGATGGGGCTTGTAGCAGAAGGCAAGCACAGGCTCAAGGCTGGAACCCCT
 15 GGTGTCTCTCCAGCGTGGTGGAAATTGGGTTAGGAGTACGGAGATGGAGATTGGCTCCCACTCTCTCC
 TATCTTAAAGGCCCACTGGCATTAAAGTGTCTATCCAAAG (SEQ ID NO: 6682)

>gi|414668|emb|X75500.1|HSMTP H.sapiens mRNA for microsomal triglyceride
 20 transfer protein
 TGCAGTTGAGGATTGCTGGTCAATATGATTCCTTCTGCTGTGCTTTTCTCTGCTTCATTTCCTCATATT
 CAGCTCTGTTTAAAGGTCACACAACCTGGTCTCTCAITAAATAATGACCGGCTGTACACAGCTCAGCTATCTC
 CACTGAAGATTCCTCTGTGATCGGGGCAAGGAAAACTGCAAGACAGGTGGGCTACCGGATTTCTCTCCAC
 GTGGATGTGGCTTACTATGGAGGAATCCGTGATGGTGTATGATGACAGTGTGATCCAAATACGATGGAAGG
 25 ATGTAAATGTTGAAAATGTAATCAGCAGAGAGGAGAGAAAGAGACTTCAAGGAGAAAAGGCCATCTAA
 AATAGTGGGAAAGGAAAACCTTGGAAAGCTCTGCAAAAGACCTACGCTCCTTCATCTAATCCATGGAAGGT
 AAAGAGTCTTACTCATATCAAAATGAGGCAAGTGGCCATAGAAAATATCAAGAGAGGTCTGGCTAGCCTAT
 TTCAGACACAGTTAAGCTCTGGAACACCAATGAGGTAGATATCTCTGGAATTTGTAAGGTGACCTACCA
 GGCTCATCAAGCAAAAGTGATCAAAATTAAGGCTTGGATTCTATGCAAAAATAGCAGAGGTCTGGATTATAG
 30 ACCCCAAATCAGGCTCTTGGGTGTCTAGTTCAAAAGCTACATCTGTCAACCACTATAAGATAGAAGACAGCT
 TTGTTATAGCTGTGCTTGTGTAAGAAAACACACAATTTTGGACTGAATTTCTCAACCACTTAAAGGGAA
 AATAGTATCGAAGCAGAAAATTAGAGCTGAAGACAAACCGAAGCAGGGCCAGAGTTGATGTCTGGAAGACG
 GCTGAGGCCATAATCAAAGCAGTTGATTCAAAGTACACGGCCATTCCCAATTGTGGGGCAGGTCTTCCAGA
 GCCACTGTAAAGGATGTCTTCTCTCTCGGAGCTCTGGCGGTCCACAGGAAATACCTGCGAGCTTGACA
 35 CTTTCCAAAGCTGAGGCTGTCAAGAACTCTCTGGCTTCATCTCAGCACCTCAGGACTGCGAAGAAAGAA
 GAGATCTTCAAACTACTAAAGATGGAATAAGGAAGTATTACCTCAGCTGTGGATGCTGTCACTCTGT
 CTCAGACCTCAGACTCATTAGAAGCCATTTTGGACTTTTGGATTTCAAAAGTGACAGCAGCATATCTCT
 CCAGGAGAGGTTTCTCATGCTGTGGATTGCTTCTCATCCCAATGAAGAACTCTGAGAGCCCTCATCT
 AGTAAGTTCAAAGTTCTTATGTGTAGCAGTGACATCAGAGAAACCTGTTATGATCATCACTGGGACACTTG
 40 TCMGAAGTGTGTGTCAGAAATGAAGGCTGCAAACTCAAAGCAGTGTGGAAGCTAAGAAATTAATCTTGGG
 AGGACTGTAAAAGACAGAGAAAAGAGGACACCAAGGATGTATCTGCTGGCTTTCGGAAGATGCCCTGCTGT
 CCAGAGGCCATCCCAAGCTCTCTGAAATGTAGCAGAGCAGGAGAGGGGCCATTCAGCACTGCGTACCA
 CTGCTCTCCAGAGATATGATCTCCCTTTCATTAATGATGAGGTGAAGAAAGACCTTAAACAGAAATATACCA
 CCAAAAACCGTAAAGTCTATGAAAGACGTGTGCGCACTGCTGCACTGTCTATCATTTTAAATTAACATCCA
 45 TCTCATAGGACGTCAAGAACACTGCTGTCTATTGTGGGAGCTTCCCCAAGAAATGAATTAATACATGC
 TCGCCATGTCTCAAGACATCTACGTTTGAATGCTCTGACAGAAATGTCCGTCGAGTCTCTGAAGGA
 AATGTCGCTCACAAATTATGACCGTTCTCCAGGAGTGGATCTTCTCTGCTCATACCTGATGATGATGTC
 CGTAGTCCCGCTTCGCGCATCTACTACAGCCTAGACATCTCTACTCTGGGTCTGCGGATTTCAAGAGAA
 GTAACCTGAACATCTTTCAGTACATTGGGAAGCTGTGTTCTTACCGTAGCCAGGTGGTTATTGAAGCCCA
 50 AGGACTGGAAGGCTTAATCGACGCCACCCCTGACGAGGGGGAGGAGAACCCTTGACTCTATGCTGGTATG
 TCAGCACTCTCTTGTGATGTCAGCTCAGACCTGTCACTTTTCAACGAGATACAGTGAATTTGATGTCCA
 AAATGCTGTGTCAGACTCTGGCCACCTATCAGTGTGGTGAAGGACATTTCTGCTCAATGATCATTTCTCA
 GGAATCTCAGTTTCAATCTGGAATAAAGCCAAATATAGAGTCCAGGCTGGTCTAGCTATTTGATATTCTCA
 GGTGCAATGGATTTAGCTTGTGTATCGTGAGTCTAAACCCCGATGAAAAATAGGGTGACTGTGTGTA
 55 TAAACCTACATCACTACAGTGAATCTCTCTTTTGTGAAGCTGGCTGGAACACAGTACAGAAACAGAGAC
 AGGCTTGGAGTTTATCTCCACAGTGCAGTTTCTCAGTACCCATCTTAGTTTGCATGCAGATGGAACAAG

5 GATGAAGCTCCATTCAGGCAATTTGAGAAAAAGTACGAAAGGCTGTCCACAGGCAGAGGTTATGTCTCTC
 AGAAAAAGAAAAAGAGGTATTAGCAGGATGTGAATTCGCCCTCCATCAAGAGAACTCAGAGATGTGCGAA
 AGTGGTGTGGCCCTCAGCCGGATAGTACTTCCAGCGGATGGTTTGAAACTGACCTGTGTGATTTTATC
 TTGGAATTTGTCTCCCGAAGGGACACAATGTGGCATGACTAAGTACTTGCTCTCTGAGAGCAGACGGTT
 10 TACNATTTTACCTGTATTTAAGATTTTGTAAAAGCTACAAAAAATGCGAGTTTGTATCAAAATTTGGGTAT
 TATGCGATGTGCTACCCACAGCGCTCATTTTGAATCATATGTGACGCTTTCAACAAGCTTTCTAGTTTATC
 TTAATCCTCTCAAAATCTCATTTTGGTACAGTCAGAAATAGTTATCTCTAAGAGGAACTAGTGTGTGTGTT
 AAAAAACAAAAAAGGCTAAGACACAAACCAATTTTCTCTTTTGTGAGTGTGGGGCCACGGGAG
 15 AAAAAAAGAAAAAGTAAAGACACAAACCAATTTTCTCTTTTGTGAGTGTGGGGCCACGGGAG
 AAGGGACAAGGCTTTTAAAGACTTGTAGCCAACTTCAAGAATTAATATTATGTCTGTATTGTTA
 GTTTTAAAGCCTTAAGGTAGAGGCACATAGAAATACATC
 (SEQ ID NO:6683)

>gi|1217638|emb|X91148.1|HSMITP H.sapiens mRNA for microsomal triglyceride
 15 transfer protein
 TGCAGTTGAGGATTGCTGGTCAATATGATTTCTTCTGCTGTGCTTTTCTCTGCTTCATTTCTCATAT
 CAGCTTCTGTTTAAAGGTCACACAATCGTCTCTCATTAAATATGACCGGCTGTACAGCTCAGTACTC
 CACTGAAGTTCTTCTTGATCGGGCAAGGAAAACTGCAAGACAGCGTGGGCTTACCGCATTTCCTCCAAAC
 20 GTGGATGTGGCCCTATCTATGAGGAATCTGTATGGTATGATGACAGGTGATCCAAATACGATGAAGG
 ATGTAAATGTTTGAATTTGAAATCAGCAGAGAGGAGAGAGAGCATCTTCAAGAGAAAGGCCATCTAA
 AATAATTTGGAAAGCGAAATCTTGAAGCTCTGCAAGACCTACGCTCTTCACTCTAATCAAGTGAAGGCT
 AAAGAGTTCTACTCATCATCAAAATGAGGCAAGTGGCCATAGAAATATCAAGAGAGGTCTGGCTAGCCAT
 25 TTCAGACAGAGTTAAGCTCTGGAACCAACCAATGAGGTAGATATCTCTGAAATTTGTAAGTGAAGTACCA
 GGCTCATCAGAGCAAGTGTACAAATTAAGGCCCTGGATTCATGCAAAATAGCGAGGTCTGGATTATGCA
 ACCCAAAATCAGGTCTTGGGTGTCAAGTCTCAAAAGCTACATCTGTCAACCTATAAGATAGAGACAGCT
 30 TTGTTATAGCTGTGCTTCTGTAAGAAACACACAATTTTGGACTGAATTTCTCACAACCATTAAGGGAA
 AATAGTATCGAAGCAGAAATTAGAGCTGAAGACACCGAAGCAGGCGCAAGATGTATGTCTCGAAGAC
 GTGCAAGCTAATCAAGCAGTTGATCTCAAGATACACGGCCATCCCAATTTGGGGCAGGCTCTCCAGA
 GCCACTGTAAAGATGTCTCTCTCTCGGAGCTCTGGCGGTCCACAGGAAATACCTGCAGCCTGACAA
 35 CCTTCCAAGGCTGAGGCTGTGAGAACTTCTCGGCCCTTCATTGACAGCTCAGGACTCGGAGAGAAAGAA
 GAGATCCTTCAAACTATAAGATGGAATAAGGAAGTATTACCTCAGCTGGTGGATGCTGTCACTCTGT
 CTGAGCTCTCAGACTATTAGAACCAATTTTGGACTTTTGGATTTCAAAAGTGACAGCAGCATATCCT
 CCAGAGAGGTTTCTCTATGCTGTGGATTGTCTTCTATCCCAATGAAGAACTCTGAGAGCCCTCAT
 40 AATAAGTTCAAAGTTCTATTTGGTAGCAGTGACATCAGAGAACTGTTATGATCATCTGAGGACATCT
 TCAGAAGTTGTGTGAGAATGAAGCTGCAAACTCAAAGCAGTATGGAAGCTTAAGAAGTTAATCCTGGG
 AGGACTTTGAAAAGGACAGAGAAAGAGGACACCAAGATGTATCTGCTGGCTTGAAGAATGCTCGCTT
 CCAGAAGGCTCCCAAGCTCTTGAAGTATGCAAGACAGGAGAGGGCCCATCAGCCACTGGCTACCA
 45 CTGCTCTCCAGAGATATGATGCTCCCTTCTATACTGATGAGGTGAAGAAGCTTTAAACAGATATACCC
 ACCAAAACCGTAAAGTTTATGAAGAACTGTGCGCACTGCTGAGCTGTCTATTTTAAATAACATTC
 ATCTCATATGAGCTCAAGAACATCTGCTGTCTATTGTGGGAGCTTCCCAAGAAATTTGCTCGTGAATCTGAAGG
 50 CTGCGCAATGTGTTCAAGCATCTTACGTTTGAAGTCTGCTGAGCAAAATTTGCTCGTGAATCTGAAGG
 AATGTGTCTCACAATATGACCGTTTCTCAGGAGTGGATCTTCTCTGCTTACACTGGCTACATAGA
 ACGTATGCTCCCGTTCGGCATCTACTTACAGCCTAGACATCTCTACTCGGGTTCTGGCATTTTAGAGGA
 AGTAACTGAACATCTTTCAGTACATTGGGAGGCTGGTCTTCAAGGTAGCAGGTGGTATTGAAGCCC
 55 AAGGACTGGAAGCTTTAATCGACCAACCCCTGACAGGGGGAGAGGAACCTTGAATCTTATGAGTGTAT
 GTGACGCTCTCTTGTATGTTTGTGCTCAGACTGTGACCTTTTCAACGATACAGTGAATTTGATGTCC
 AAAATGCTGTCTGAGCTCTGGGACCTATCAGTGTGGTGAAGGACTTATCTGCTTAATAGATCATCTCT
 AGGAATCTCAGTTACATCTGTGATCAAAAGCCAAATAGAGGTCCAGGTTGGTCTAGGTATTGATTTTAT
 60 ATAAACCCTGACATCAGGTGACCTCTCTTTTGTGAAGCTGGCTGGAATCCCATCTAGTTTGAAGTGAAGCA
 CAGGCTTGGAGTTTATCTCCACAGTGCAGTTTCTCAGTACCAAGGCTGTCCACAGGCAGAGTTATGTCTCT
 GGTGAGGCTCATTCAGGCAATTTGAGAAAAAGTACGAAAGGCTGTCCACAGGCAGAGTTATGTCTCT
 CAGAAAAAGAAAGCGATATTAGCAGGATGTGAATTTCCCGCTCCATCAAGAGCACTCAGAGTGTGCA
 65 AAGTGTGTGTTTCCCTCTCAGCGGATAGTACTTCCAGCGGATGGTTTGAAGTACAGCTGTGATTTTGA
 CTTGTAAATTTGTCTCCCGAAGGACACAAATGTGGCATGACTAAGTACTTGCTCTCTGAGAGCAGCGT
 TTACATATTTACCTGTATTTAAGATTTTGTAAAAAGTACAAAACTGCAATTTGTATCAAAATTTGGGT
 ATATCAGTATGCTACCCACAGGCTCATTTTGAATCATCATGTGACGCTTCAACAACGCTTTCTAGTTTA

CTTATACCTCTCTCAAAATCTCATTTGGTACAGTCAGAAATAGTTATTCTCTAAGAGGAAAACTAGTSTTTGT
 TAAAAACAAAATTAFAAACAAAACACACAGAGGAAACCAATTTTGTCTCAACAATTTTGTGATCAATGT
 ATATGTAGCTCTCTGTATAGGACTTCTTAAAGCATGACGGGAAAACCAACACAGCTTCCCTAATCAGGAAAAA
 AAAAAAAAAGAAAGATTAAGACACAAACAAACCAATTTTCTCTCTTTTGTGGAGTTGGGGGCCCGGAGG
 5 AGAAGGGACAAGCTCTTTAAAAGACTTGTGTAGCCAACTTCAAGAAATTAATATTGTCTCTGTATTATTGT
 TAGTTTTTAAGCCTTAAGTGTAGAAGGCATAGAAATAACATCTCATCTTTCGTGTCAGCAATTTAGTGTAG
 GTTGTTCACAAAGGACATTCAGGTCTCTACCTCCAGCCCTGCAGAAAATATTGGACCTAGACACAGAGCAATC
 AGGAAAAATTAATTTCAGAAACTCCATTTGATTTTCTTTTGTCTGTCTTTTGTGAGACTGTAAATAGGT
 10 ACATCTCTCCTTAAGGACATCTCTCATTTATCTCACCTTTTGGGGGTGAGAGCTCTAGTTTCATTTAACT
 GTACTCTGCACAAATAGCTAGGATGACTAAGAGAAACATTCGTCTCAAGAAACTGTGTGGATTGTGGATTTCCTAA
 AATATGAAATAAGGAGAAAAATGTTTTTATTGTATGAATTAAGATCCATGTGTGAACATTTGCAAAATA
 TTTTATTAAACAGATGTGGTGATAAACCCAAAAACAAATGACAGGTGCTTATTTTTCCACTAAACACAGA
 CACATGAAATGAAGTTTAGCTAGCCCACTATTGTGTGTAATTGAAAACGAAGGTGTGATAAAATAAATA
 TGTAGAAATCAAAAAAATAAAAAA (SEQ ID NO: 6684)

>gi|21361125|ref|NM_001467.2| Homo sapiens glucose-6-phosphatase, transport
 (glucose-6-phosphate) protein 1 (G6PT1), mRNA
 20 GGACACGAGAGGACCCAGGAGCGCTGCTCCCTGACCACGAGCAGAGACCCCTTCTATCGCGCCAGTCCCTG
 TGGTCTTCGCGACCTCTCCAGCTCTGTCACCCCGGCCCGCTGTTTCCAGCCGCGACAGTAGCGTGTCTCT
 GGGTAGCGTGTAGAGCCACGGGGCTGTAGCAGGTGCCACGAGCCCGCGCTCTTTCGCGCCGCGCGCTCT
 TCTCTCTTCCCGACCGCGCGCTGCGCCCTCCCTTACGAGCTGTAGCTCTGCGCTGTGAGCGGCGGAGCGG
 25 CTCCGGGAGAAATAGTTCGCTGCGCGCTCTGCGCTGTGCGCTTCTGCTCTCAGTTGCTTTTCCCAAT
 TGAGCGGAAAAGCCGGGCGATTTGCGGGGCGCTTGGCGGCGAGCTGTGCGCTGTGAGCGGCGGAGCCG
 CCGGGGACCTTCCCGCCACGAGCTGCCAGTCTGCTGTGCGGGCCCTGCCAACGCTCCACAGGAC
 ACTGTGGTCCCTTTGGAGCCTCCCGAGCTTATGATTTGTGAGAGCGGCGATTAAGAGGAGCTTGGAG
 GCTGGGTGGAGAGGGAGCAGAAAAAACCAACTCAGCAGATCTGGGAACGTGTGAGAGCGGCAAGCAGGA
 ACTGTGGTCAGAGGCTGTGCGCTTGTGCTGTAGGCGCTGCTCTTTTCTACCATGGCAGCCAGGGCTAT
 GGCATTATTGCGACTGTGATCTTCTCAGCCATGTTTGGGGCTACAGCTGTGATTACTTCAATGCAAGA
 30 CCTTCTCTCTTGTGATGCCATCAITGGTGGAGAGATCCCTTTGGACAAGGATGATTGGGGTTCCATCAC
 CAGCAGCCAGTGTGCGCAGCTTATGCTATCAGCAAGTTTGTAGTGGGGTGTCTGTGACACAGATGAGTGT
 CGCTGTGCTCTTCTCTTGTGGCTGTCTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT
 TACCTGTCTTGTGCTGCGCTTGTGCTTCTTAATGGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT
 35 GTGCTGTGCGGAAGTGTGTGAGCCATCTCAGTTTGGCACTTGTGTGGCTGTGCTGTGCTGTGCTGTGCTGTGCT
 CTGCTGTGGAAGGCTGGGCGCTATCTGGCAACCATCTTGCCAGAGCTACAGCTGTGCGCAGCAGCTGTG
 CCTATCTGGGGCACTGT
 TGTGTGACTCCGCACTGTGAGCCCATGCGCTCTGAGGGCAAGAGGGCTCTTGAAGAGGAGAGACCACT
 CTGAGGAGCTGCTGCTGTGCCCTTACCTGTGGTGTCTTCCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
 40 AGACTGT
 CTCTCATAGTGTGCTGTGAGGTGTGGGGCTGTGTAGGCGCATGTGAGCTGTGTGTGTGTGTGTGTGTGT
 GGCATGGCAAGGGCGGAGCTGTGCAACTACGGGAACCTCGCCATGGCTGTGTGTGTGTGTGTGTGTGTGT
 GCATGACAGTGTCCATGTACCTCTTCCGGGTAAACAGTGTGACAGTGTGTGTGTGTGTGTGTGTGTGTGT
 45 ATTGGAGCTGTATTGTGTTTCTCTCGTATGCCCCATGTGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGT
 GCGCTTCTGCTTGTGTCAGCCATGTGCAAGCACTACAGTTGTGAGCAGCAGCTTGTGGGTGTGTGTGTGT
 TGT
 AAGGCTGT
 50 GGGAAAAAGAGGGGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT
 CCCGGTGTGCGCTGTGAGATTATCAGTGTGCTGTGAGGTCCAGCTTCCGTGATCTTGTGCTGTATTAAAA
 GATAACCTTGTGCTGTGACTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGT
 CAGGAAGGCTCTGTGACCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
 55 GAGTGAGGCTATCCCTGTGCTGTGAGGCGCTATGACCTTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGT
 AGCATGGGTGTGCTGT
 ACCACTAGGCGCAGTTGTGATAGTGTGGGAAGAAAGGTGACTTGTATAGAGATTAAACATAGATTGTG
 ATACTGAAAAAATAAAAAAATAAAAAA (SEQ ID NO: 6685)

gi|4503130|ref|NM_001904.1| Homo sapiens catenin (cadherin-associated protein),
beta 1, 88kDa (CTNBN1), mRNA

5 AAGCCTCTCGGTCTGTGGCAGCAGCGTTGGCCCGGCCCGGGAGCGGAGCGGAGGGGAGCGGAGACGG
AGGAAGGTCCTAGGAGCAGCTTCAGTCCCGCCGACCGCCAGCAGGTCTGAGGACAGGTCTCGGACTCCCG
CGCGGGGAGGAGCCTGTTCCTCTGAGGGTATTTGAAGTATACCATGCTGTTTGAAGTATACAGCGTG
GACAAATGGCTACTCAGCTGATTTGATGGAGTTGGACATGGCCATGGAACAGACAGAAAGCGGCTGTT
AGTCACTGGCAGCAACAGCTTTACCTGGACTCTGGAATCCATTCTGGTGCCACTACCAAGCTCTCTCTC
10 TGAGTGGTAAAGCAATCCTGAGGAAGAGGATGTGGATACCTCCCAAGTCTCTGTATGAGTGGGAACAGGG
ATTTTCTCAGTCTCTACCTCAAGAAACAAGTACTGATATTGATGGACAGTATCAATGACTCGAGCTCAG
AGGGTACGAGCTGCTATGTTCCCTGAGACATTAGATGAGGCGATGCAGATCCCATCTACACAGTTTGAAT
CTGCTCATCCCACTAATGTTCCAGCGTTTGGCTGAACCATCACAGATGCTGAAACATGCAGTTGTAAACTT
GATTAACTATCAGATAGTATGCAGAACTTGGCCACGTCGAATCCCTGAACTGACAAAACTGCTAAATGAC
GAGGACCAAGTGGTGGTTAATAAGGCTGCAGTTATGTGTCATCAGCTTTCTAAAAGGAGAGCTCCAGAC
15 ACGCTATCATGCGTTCTCTCCAGATGGTGTCTGCTATTGTACGTACCATGCAGAATACAAATGATGTAGA
AACAGCTCGTTGTATACCGCTGGGACCTTGCATAACCTTCCCATCATCGTGAGGGCTTACTGCGCAATCTT
AAGTCTGGAGGCATTCCTGCCCTGGTGAATAAGTCTGGTTACCAAGTGGATTCTGTGTTGTTTATGCCA
TTACAACTCTCCCAACCTTTTATTACATCAAGAGGAGCTAAAATGGCAGTGGCTTTAGCTGGGGCT
20 GCAGAAAATGGCTTGCCTTGTCTCAACAAAACAATGTGTAATCTTGGCTATTACGACAGAGCTGCCTCAA
ATATTAGTCTATGGCAACCAAGAAAGCAAGCTCATCATACTGGCTAGTGGTGAGCCCAAGCTTTAGTAA
ATATTAATGAGGACCTACTACTTACGAAAACTACTCTGGACACAAGCAGAGTGTCTGAAGGTGCTATCTGT
CTGCTCTAGTAATPAGCCGGCTATTGTAGAAGCTGTGGAAATGCAAGCTTTAGGAGCTTACCTGCACAGAT
CCAAGTCAACGCTCTTGTTCAGAACTGCTTTGGACTCTCAGGAATCTTTCAGATGCTGCAACTTAAACAGG
AAGGATGGAAGGTCTCTCTGGGACTCTTGTTCAGCTTCTGGGTTCAGATGAATATAATGTGGCTCACCTG
25 TGCAGCTGGAATCTTTCTAACTCAGTTGCAATAATTATAAGAACAGATGATGGCTGCCAAGTGGGT
GGTATAGAGGCTCTTGTGCGTACTGTCTCTCGGCTGGTGACAGGGAAGACATCACTGAGCTTGCCATCT
GTGCTCTCTGCTATCTGACGACCGGACACCAAGAGCAGAGATGGCCAGAAATGCAAGTTGCGCTTCACTA
TGGACTACCAATTTGGTGTAAAGCTCTTACACCCACCATCCCACTGGCCCTTGATAAAGGCTACTGTTGGA
TTGATCGAAATCTTGCCCTTTGTCCCGCAAAATCATGACCTTTGGTGAGCAGGCTGCCATTCCACGAC
30 TAGTTCAAGTTGCTTGTCTGTCACATCAGGATACCCAGCGCGTACGTCCATGGGTGGGACACAGCAGCA
ATTTGTGAGGGGGTCCGCGATGGAAGAAATAGTTGAAGTTGTACCGGAGCCCTTCACATCTAGCTCGG
GATGTTCAACAACCGAATTGTTATCAGAGGACTAAATACCAATCCATTGTTTGTGACGCTGCTTTATTCTC
CCATTGAAACATCCAAAGAGTAGCTGCAGGGGTCTCTGTGAACCTTGCTCAGGACACAGGAAGCTGCAGA
AGCTATTGAAGCTGAGGGAGCCACAGCTCCTCTGACAGAGTTACTTCACTTAGGAATGAAGGTGGGCG
35 ACATATGACAGCTGCTGTTTGTTCGGAATGCTGAGGACAGGCCACAAGATTACAAGAACCGCTTTTTCAG
TTGAGCTGACAGCTCTCTCTCAGAACAGAGCCAAATGGCTTGAATGAGACTGCTGATCTTGGACTTGA
TATTGGTGCCAGGGAAGACCCCTTGGAATATCGCCAGGATGATCTAGCTATGTTCTTTTCACTCTGGT
GGATATGGCCAGAGTACCTTGGGTATGGACCCCATGATGGAACATGAGATGGGTGGCCACCCCTGGGTG
CTGACTATCCAGTTGATGGGCTGCCAGATCTGGGCTATGCCAGGACCTCATGGATGGGTGCGCTCCAGG
40 TGACAGCAATCAGCTGGCTGGTTTGATATGACTGATTAATCATCTTTAGCTGTATTGTCTGAACTTG
CATTTGATTGGCTGTAGAGTTGCTGAGAGGGCTCGAGGGGTGGGCTGATCTGCAAAAGTGCTGAC
ACACTAAACAGCTGAGTTTCCATAGGGAACAATTAAGATAAATCTTTTGTCTGGTCCCTTTTGGTGCAG
GGAATACAAATACAAATGGATTTTGGAGTGAATCAAGAGTGAAGATGCACAGGAATGGATCAACAGA
TGGAAATTAGCAAAACCTAGCCTTGCTTGTATAAATTTTTTTTTTTTTTTTTTAAAGATATCTGTAATG
45 GTACTGACTTGTGCTTCTTGAAGTAGCTCTTTTTTTTTTTTTTTTTTTTTTGTGCACTAGCTGTTT
TTTTAGTCTCTCGTAGTGTTAAGTTATAGTGAATACTGCTACAGCAATTTCTAATTTTTTANGAATTTGAGT
AATGGTGTAGACCACTAATTAATCATAATCACTCAATTAATTTGTAATCTGAATAAAGTGTAACAATGT
TGTAAGCTTTTTGTATAAATAGACAAATAGAAAATGGTCAATTAGTTTCCCTTTTAAATATGCTTAAAA
TAAGCGGTGAGATCTATTTCATGTTTGTGATCAAAACTATTGGGATATGTATGGGTAGGGTAAATCAG
50 TAAGAGGTGTTATTGGAACTTGTTTGGACAGTTTACAGTTTGCCCTTTTATCCCAAGTGTGTGTAAC
CTGCTGTGATACGATCTTCAAGAGAAATCGCGTTATAAAAATGGTTGCAATTAACCTTTAATTCA
TT (SEQ ID NO:6686)

gi|18104977|ref|NM_002827.2| Homo sapiens protein tyrosine phosphatase, non-receptor type 1 (PTPN1), mRNA

5 GTGATGCGTAGTTCGGGCTGCCGGTTGACATGAAGAAGCAGCAGCGGCTAGGCGCGCGGTAGCTGCAGGG
GTCGGGGATTGCAGCGGGCCTCGGGGCTAAGAGCGCGCAGCGCGCTAGAGCGGCAGACGGCGCAGTGGCG
CGAAGAAGAGGCGCAGCAGCGCCCTGGCCCGTCAATGGAGATGGAAAAGAGATTGAGCAGATTCGACAAG
10 TCCGGGAGCTGGCGGCCATTACACAGGATATCCGACATGAAGCCAGTGACTCTCCCATGTAGAGTGGCCA
AGCTTCTTCAAGACAAAACCGAAATAGGTACAGAGACGTCACTCCCTTTGACCATAGTCGGATTAAACT
ACATCAAGAGATATTAAGTATATCAACGCTAGTTTGTATAAAATGGAAGAAGCCCAAAGGAGTTACATT
CTTACCCAGGCGCCCTTTGGCTAACACATGCGGTACACTTTTGGAGATGGCGACACATCTGGCCACAAAAGA
15 GTGTGCTCATGCTCAACAGAGTGATGGAGAAAAGGTTGCTTAAATGGCGACACATCTGGCCACAAAAGA
AGAAAAGAGATGATCTTTGAAGACACAAATTTGAAATTAACATTGATCTCTGAAGATATCAAGCTCATAT
TATACAGTCGCAGCAGCTAGAATTGGAAAACCTTCAACCCCAAGAACTCGAGAGATCTTCACTTTCCACT
ATACCACATGGCCCTGACTTTGGAGTCCCTGAATCAACAGCGCTCATCTTGAAGCTTTCTTTCAAGAGTCGG
AGAGTCAGGCTCACTCAGCCCGGAGCAGCGGCCCGTTGTGGTGCACTGCAGTCGACGAGTACGTCGGCAGGTTCT
20 GGAACCTTCTGCTGCTGATACCTGCCCTCTTGCTGATGGACAAGAGGAAAGACCCCTCTCTCCGTTGATA
TCAGAAGAGTCTGTTAGAAAATGAGGAAGTTTCGGATGGGGCTGATCCAGACAGCGCAGCTCGGCTT
CTCTCACTGGCTGTGATCGAAGGTGCCAAATTCATCATGGGGGACTCTCTCCGTGAGGATCAGTGGGAAG
GAGCTTTTCCCAAGAGGACTGGGAGCCCCACCCGAGCATATCCCCACCTCCCGCGCACCCCAAGCA
25 TCTCTGAGGCACACATGGGAAATGAGGAGAGTTCTTCCCAAATCACCAGTGGGTGAAGGAAGAGACCCA
GGAGGATAAAGACTGCCCATCAAGGAAGAAAAGGAAGCCCTTAAATGCCGACCCCTACGGCATCGAA
AGCATGAGTCAAGACACTGAAAGTTAGAAGTGGGTCGTGGGGGGAAGTCTTCGAGGTCCGACCGTCCCT
CCCCAGCCAAAGGGGAGCCGCTCACTGCCGAGAGAGCAGGACCATGCACTGAGTTACTGGAAAGCCCTT
CCTGGTCAACATGTGCTGGCTACGGTCTCTCAACGCGCGCGCTTACCTCTGCTACAGGTTCTGTGTTCAAC
AGCAACACATAGGCTGACCTCCCTCCACTCCACTCCACCCACTGTCGCGCTCTGCGCGCAGAGGCCACG
30 CCGGACTAGCAGGCATCCGCGGTAGGTAAAGGCGCGCGGACCGCTAGAGAGCCGGGCCCCGGAGCGAC
GTTGGTTTCGCACTAAACCCATCTTCCCGGATGTGTGTCTCAACCCCTCATCTTTATCTTTTTGGCCCC
TTCCACTTTGAGTACCAATCCCAAGGCCATTTTTTGGAGGAGTGAAGAAGAGATCTTTCGCGCGG
CAGAGGGAAGGGGCTACACCCGCTTGGGGCTCGCGCCACCCAGGGCTCCCTCTCGAGCATCTCCGAGG
GGGCGGCACGCCAACAGCCCCCCTTGAATCTGACGGGAGCACTCTCCACTCATATTTTATTAACA
35 ATTTTTCCTCCCAAGGCATCCATAGTGCACTAGCACTTTCTTGAACCAATAGTATTAATAAATTTTGA
TGTCAAGCTTGCATCAAGGGCTTTATCAAAAAGTACATAA TAAATCCTCAGGTAGTATCGGGAATGGAA
GGCTTTGGCATGGGCTGCTGCGTCAGACAGTACTGGGAAGGAGGACGGTTGTGAAGCAGTTGTTAATTA
GTGATATTTGTGGTAACTGAGAAGATAGAACAATGCTATAATATATAATGAACACGTTGGGTATTAATA
AGAAACATGATGTAGAGATTACTTTGTCCCGCTTATCTCTCCCTGTTATCTGCTAGATCTAGTTCTCAA
40 TCACCTGCTCCCGCTGTGATTAGAAATGCATGTAAGGTCTCTTGTGCTCGTATGAAAATTTGAGTGTG
AAATGAGAAATCTTATCTCTGCTAATAATGTGCCCATGTCCAAGTCCAACTGCTCTGGATGACCT
GATCATTACATGCTGTGTTGCTTCAAGCTTGTTGCTGAAGTCATTGTGCTCAGCAATGTGAGTGCAGTTT
TCCAGGAATAGGCATTTGCTTAATCTTGGCATGACACTTGTGACTTCTCGTGAAGCGCCAGGCTGCTC
45 TCCGATCAGCAGGCTTTGCTGTAACTCAGACATTTCAAGGGTATGGGAAGCATATTCACACTCACTGCT
CTCGGACATGATTAGGAAGCAGGGAACCCCGCGCCCACTTTGGGATGAGACTCTCCGCAATCCCA
AGTCAACATCTCTTTGAGCAGACCGTGATTTGGAAGAGAGGACCTCTGGAAACCACTACTCTTTGAAA
CAGCTCTGGGTACGGTCTTTTAGGCAGCTTCCCGCCGCTCTCTGTGCTCGTATGAAAATTTGAGTGTG
CGTCTGCCCACTCTCAACCTCTGTGGGGCTGTAGTGTGCTCAGCACTCTCTCTGCAAGGGAAGCTGAAG
50 ACCTTCACATTAAGTGGCTTTTAACTGAAAACACGCGAGCTGTAGCTCCCGAGCTACTCTCTTGC
GCATTTTCACATTTTGGCTTTCTGCTGTGTAAGCCAGTACAGAGAAATTTCTGTGTGGGAACATCTCGAG
GTGTACCCCTCCAGAGCTATGTGTGAGGTGTGGATAAGGCTTAGGTGCCAGGCTGTGAAGCTCTGAGCTG
GGCTTGTGTTTAAAGCTCTGTATATGTATGTAGTAGTTTGGGTGTGTATATAGTAGACTTTCAAAA
TGGACGTATCTGGTTTAACTCTTACTCTTGGAGAGCAGCTGGCTCTCCACTGTGTACACATTATGTAGT
AGAGGTAGCAGCTGCTCTGCTATATGCCCTTAAGCCAAATATTACTATCAGGTCAATATTTTACAAT
55 GGCCATGGAATAAACCAATTTTACAAA (SEQ ID NO: 6687)

gi|12831192|gb|AF333324.1| Hepatitis C virus type 1b polyprotein mRNA, complete cds

55 GCCAGCCCCCGATTGGGGGCGACACTCCACCATAGATCACTCCCTGTGAGGAACACTGTCTTTCACGCA
GAAAGCGTCTAGCCATGGCGTTAGTATGAGTGTCTGTGAGCTCCAGGACCCCCCTCCGGGAGAGCCA

TAGTGGTCTGCGGAACCGGTGAGTACACCGGAATTGCCAGGACGACGGGTCCTTTCTTGGATCAACCCG
 CTCATTCGCTCGGAGATTGGGCGTGCCCCGCGAGACTGTAGCCGAGTAGTGTGGGTGGCGAAGGCC
 TTGTGGTATGCTGTATAGGGTCTTGCGAGTGCCCGGGAGGGTCTCGTAGACGCGTGCTCATGTAGACACA
 AATCTCAATAACTCAAGAGAAAAACAAACGTAAACACCAACCGCCGCCACAGAGCTTAAGTTCGCCGGCG
 5 GTGTGCAGATCTGTGGTGGAGTTTACTGTGTCGCGCGAGGGGCCCAAGTGTGGGTGTGCGCGCGACTAG
 GAAGACTTCCGAGCGGTGCGCAACTCGTGGAAAGCGCAACCTATCCCCAAGGCTCGCCGCCCGAGGGT
 AGGACCTGGGCTCAGCCCGGTACCTTTGGCCCTCTATGGCAACGAGSGTAGTGGGGTGGGCGAGATGGC
 TCTGTCAACCCTGTGCTCTCGGCTAGTTGGGGCCCAAGACACCCCGCGCTAGCTCGGTATTTGGG
 TAAGGTCATCGATACTTACATGCGGCTTCCGCGACCTCATGGGGTACATTTCGCTGTGCGGCGCCGCC
 10 CTAGAGGGCGTCGCCAGGGCCCTGGCGCATGGCGTCCGGTTCTGGAGAGCGCGGTGAATATGCAACAG
 GGAATCTGCCCGGTGCTCTTCTCTATCTTCTCTTAGCTTTGCTGTGTGACCACTCCAGCTCTC
 CGCTTACGAGGTGCGCAAGTGTCCGGGATATACCATGTACAGAACGACTGCTCCAACCTCAAGTATTGTG
 TATGAGGCGAGCGGACATGATCATGACACACCCCGGGTGTGGTCCCTGCGTCCGGAGAGTAATTTCTCC
 GTTGCTGGGTAGCGCTCACTCCCAAGCTCGCGCGCAGGAACAGCAGCATCCCAACAGCAATACGACG
 15 CCACGTCGATTGCTGCTGTGGGCGCGCTGCTCTCTGTTCCGCTATGTACGTTGGGATCTCTGCGGATCC
 GTTTTCTCGTCTCCAGCTGTTACCTTCTCACTCGCCGATGAGACGATACAAGATTGCAATTGCT
 CAATCTATCCCGGCCAGTATCAGGTCAACGATGCTGGGATATGATGATGAATGCTGCTCACTCAAC
 GGCCTAGTGGTATCGCAGCTACTCCGGATCCCAAGCGGTGCGGGGCCCACTGG
 GGTGTCTAGCGGGCTTGTCTACTATTCCATGGTGGGGAATGGGCTAAGTCTGTGATGTGATGTCTAC
 20 TCTTGTGCTGGTGTGACGGGCACACCCAGTGCAGAGGGGAAGGGTAGCCTCCAGCACCCAGAGCCTCGT
 GTCTGTGCTCTCAAGAGGCCATCTCAGAAAAATCCAACCTCGTGAACACCAACGAGCTGGCACATCAAC
 AGGACCGCTCTGAATTGCAATGACTCCCTCCAACCTGGTTCATTGCTGCTGCTGCTGCTGCTGCTGCTG
 TCAACGCGTCTCGATGTCCAGAGCGCATGGCAGCTGCCGCCCATCGACAGATTCGCTCAGGGGTGGG
 25 TCCCATCTCAAGTGTGCTTCAACTCTCGAGCAGAGGCTTATTGCTGGCACTATGCAACCCCAACG
 TGCAGTATTGTACCGCGCTCGCAGGTGTGTGCCAGGTGATTGCTTCAACCCGAGTCTGTTGTGTGGT
 GGAAGACCGAGCTTCCGGAGTCCCACTATAGCTGGGGGAGAAATGAGACAGAGCTGCTGCTCACTCA
 CAACAGCGCGCCGCGCAAGGCACTGGTTTCGCTGTACATGGATGAATGACACCGGGTTCACCAAGAG
 TGGCGGGGCCCGCGTGTAACTATCGGGGGGTGGCAACCAACCTTGAATTGCCCCAGGATGTGCTTCC
 30 GAAAGACCCCGAGGCCACTTACACCAAAATCGGCTCGGCTCGGCTTGTGTGACACCTAGGTGTCTAGTTGA
 CTACCAATACAGACTTTGGCACTACCCCTGCACTATCAATTTACCATCTTCAAGCTCAGGATGTAGTG
 GGGGCGGTGGAGCACAGGCTCAACGCGCGCTGCAATTGGACCCGAGGAGCGCTGTAGCTGTGGAGACA
 GGGATAGATCAGAGCTTAGCCCGCTGCTATTGTCTACAACGAGTGGCAGGTACTGCCCTGTTCTTAC
 CACCTTACCGCTCTGCTCCACTGGATTGATCACCTCCATCAGAATATCGTGGAGCTGCAATACCTGTAC
 35 GGTGTAGGTCAGTGGTGTCTCCGTCGTAACTAAATGGGAGATGTTCTGCTGCTCTCTCTCTCTCTG
 CGAGCGCGCGCTCTGTGCTGCTTGTGGATGATGCTGCTGATAGCCAGGCTGAGCGCTTAGAGAA
 CCGTGGTGTCTCAATGGCGCTCTGTGGCCGGAGCGCATGCGCTTCTCTCTCTCTGTTCTCTGCTG
 GCGCGCTGCTACATCAAGGCAAGCTGGTCCCTGGGCGGCAATATGCTCTCTATGGCTCAGGGATGGG
 TCTCTGCTCTCTGCGCTTACCAACAGAGCTTATGCCATGGAACCGAGAGATGGCTGCTGCTGCGGAG
 40 CGCGGTCTTGTAGGCTCTGGTACTCTTGACCTTGTCAACATACTAAGGTGTTCTGCTGAGGCTCAT
 TGGTGGTTACAATATTTTATCACCAGGCGCGAGGCGCACTTGCAGTGTGGGTGCCCTCTTAATGTT
 GGGGAGGCGCGATGCCATCATCTCTTACATCGCGGTCCATCGCAGAGTAATCTTTGACATCAACA
 ACTCTGCTCGCCATCTCGGTGCTCATGGTCTCGAAGCTGCGATAACAGAGTCCGCTACTCTGTG
 CGGCTCAAGGCTCTATTCTGTCATGATGTTAGTGGGAGAGTCTGCGGGGTCTATTGTCCAAATGG
 45 CTTCTAGAGCTGGGCGCGCTGACAGGCACTTATTAACCACTCTAACCCGCTCAGGGATGGG
 CCACGCGGCGCTACAGGACTTGCCTGGCAGTGGAGCCGCTGCTCTCTCCGATGAGAGCAAGAT
 ATCACTTCGGGAGCAGACACGCGCGCTGTGGGACATCATCTTGGGTCTCCGCTCTCGCCGAGAGG
 GAAAGGAGATATCTCTGGGCGCGCGGATAGTCTTGAAGGCGGGGGTGCGACTCTCTCGCGCCATCAC
 GGCCTACTCCCAACAGAGCGCGGGCTACTTGGTTCATCACTACGCTTACGCGCGGAGCAAGAAC
 CAGGTGCGAGGAGAGGTTCAGGTGTTTCCACGCAACCAATCTTCTGCGGACCTGCTGCTCAACGG
 50 TGTGTTGACCGTTTACCATGGTGTGCTGCTCAAGACCTTAGCCGCGCCCAAGGGCCCAATCAACGAG
 GTACACTAATGTGGACAGGACCTGTCGCTGCGAGGCGCGCCCGGGGCGGTCTCTTGACACATGCT
 ACTGTGGCGACCTCAGACCTTTACTTGTGTCAGAGCATGTGACGTCATCTCGGTGCGCGCGGCG
 ACAGTAGGGGAGGCTGCTCTCCCCAGGCTGCTCTTACTTGAAGGGCTCTTCCGGTGGTCCACTGCT
 55 CTGCTCTCGGCGCGCTGTGGGCTCTTCCGGGCTGCGGTATGACCCGCGGGGTGTGGAGGCGGTG
 GACTTTGTGCTCGTAGAGTCCATGGAACATACTATGCGCTCTCCGCTCTCACGAGCACTCATCCCC
 CGGCGGTACGCGAGTCTTTCAAGTGGCCCACTACAGCTCCCATCGGACGCGCAAGAGTACTAAGGT
 GCGCGCTGCATATGTGACGCCAAGGTTACAGGTGCTGCTCTCAATCGTCCGTTGCGCGCTACTTAA
 TTTGGGGCGTATATGTTCAAGGCACAGGTATTGACCCCAACATCAGAACTGGGGTAGGACCATACCA

CAGGCGCCCCGTCACATACTTCACTTATGGCAAGTTTCTTGCCGATGGTGGTTGCTCTGGGGGCGCTTA
 TGACATCATATAATATGTGATGAGTGGCATTCAACTGACTCGACTACAATCTTGGGCATCGGCACAGTCTCTG
 GACCAGCGGAGACGGCTGGAGCGCGGCTTGTCGTGTCTGCCACCGCTACGCTCTCGGGATCGGTTCACCG
 5 TGCCACACCCAACTACGAGGAGGTGGCCCTGTCTAATATCGGAGAGATCCCTCTCTATGGCAAGCCAT
 CCCCATTGAAGCATCAGGGGGGAAGGCATCTCAITTTCTGTCAITCCCAAGAAAGATGCGACGAGCTC
 GCGCGAAGCTGTTCAGGCTCTCGGAATCAACGCTGTGGCGTATTACCGGGGGCTCGATGTGTCCGTCATAC
 CACTATCGGAGACGTCGTGTCTGTGCAACAGACGCTCTGTATACGGGCTATACGCGGCGATCTTGAAGT
 AGTGAATGCACTGTATCAACATGTGTCCACCAGACAGTCGACTTCAGCTTGGATCCCACTTCCACATTTAGAG
 ACGACGACCTGTGCTCAAGACGCAGTGTCTGCGCTCGCAGCGCGGGTATGGATCTGCGAGGCTTAGGAGAG
 10 GCATCTACAGGTTTGTGACTCCGCGGAGAACGCGCCTCGGGCATGTTGCAATTCCTCGGCTCTGTGTGAGTGT
 CTATGACGCGGGCTGTGCTTGTGTACGAGCTCACCCGCCCGAGACCTCGGTAGGTTCGGGCGCTCACTCTG
 AACACACACGAGGTTGCCGCTTTCGAGGACCACTTGGAGTTCTTGGAGAGATGTCTTCAACAGGCTCCACCC
 ACATAGATGTCACATCTTCTTCTCCAGCAACAGCAGGCGAGAGACAACTTCCCTTACCTTGGTAGCATACCA
 AGCCACGCTGTGCGCCAGGGCTCAGGCCCCCACTTCCATCATGGGATCAAATGTGGAGTGTCTCATACGG
 15 CTGAAACCTACGCTGACGCGGCCAACACCTTGTCTGTACAGGCTGGGAGCGTCCAAAATGAGGTCACCC
 TCACCCACCCCTCAACAAATATACATCATGGCATGCATGTGCGCTGACCTGGAGGTCGTCACTAGCACCTG
 GGTCTGTGTGGCGGAGTCCCTTGACGCTCTGCGCGCGTATTTGCTGACACAGGACAGTGTGTGTCATTTGTG
 GGTAGGATATCTTCTTCCGAGGCGCGCTATTGTTCCCGACAGGAGCTTCTCTACCAAGAGTTTCGATGT
 AATGTGAAGATGTGCGCCACGCACTCCCTTTACATTGAGCAGGGAATGTCAGCTCGCCGAGCAGTTTCAAGCA
 20 GAAAGCGCTCGGGTATCTGCAACAGCCACCAACAGCGGAGGCTGCTGTCTCCGTGGTGGAGTTCGAG
 TGGCGGCTCTTGAGACATTTGCGCGAAGCACATGTGGAATTTTACAGCGGGATACAGTACTTAGCAG
 GCTTATCCACTGTGCTTGGGAACCCGCAATAGCATCTTGAAGCATTCAGGCTTCTTACACGAGCC
 GCTCACCCCTCAAGATCCCTCTCTGTATTAACATCTTGGGGGGGTGGGTGGCTGCCAACTCGCCCCCCCC
 AGCGCGCTCTGCGCTTTCGTGGCGCCCGGCATCGCGGTGGCGCTTGGGACGATAGGGCTTGGGAAGG
 25 TGCTTGTGGACATCTTGGCGGGTTATGGAGCAGGATGGCGCGCTCGCTGTGGCTTTAAGGCTCATAG
 CCGCGAGATGCCCTTCCAGAGACCTTGGTCAATCTACTTCTCCATCCCTCTCTCGCGCCCTGAGT
 GTCGGGCTCGTGTGTGACGCAATCTGCTGCTCGGCAGTGGGTCCGGGAGAGGGGGCTGTGCAATGGATGA
 CCGCGCTGATAGGTTTCTGCTCGCGGGTAACTACGCTTTCGCCACGCACTATGTGCTCGAGAGCGAGCC
 30 CAGCGCGGTGTACTCAGATCTCTCCAGCTTACCATCTACAGCTGTGTAAGAGGCTCCACAGTGG
 ATTAATGAGGATGCTCCACACCGTGTTCGCGCTGTGCTTAAGGATGTATTGGGACATGATATGACGCG
 TGTGTGACTGACTTCAAGACCTGGCTCCAGTCCAAAGCTCTCGCGCAGTACCGGGATGCCCTTTTCTCTC
 GTGCCAACCGGGTCAAGGGAGTCTGCGGGGAGACGCGCATCATGCAAAACCACTTCCCATATGTGGAGCA
 35 CAGATACCCGACATGTCAAAAACGCTTCCATGAGGATCTGCGGGCTTAAGACCTGACGCAACACGTGGC
 ATGGAACATTTCCCATCAACGCATACCAACGCGGCCCTTGACACACCTCTCCAGCGCCCAACTCTTCTAG
 GGCCTGTGGCGGGTGGCCGTGAGGAGTACGTGAGGTTCACGCGGGTGGGGATTTCCACTACGTGAGT
 GAGTACGACCTGACAACTAAAGTGCCATGCGAGGTTCCGGCTCCTGAATTTCTTCTCGGAGGTGGAGCG
 GAGTGGGGTTCACAGGTACGCTCGCGGTGACGCGCTCTCTTACGCGAGGAGGTTTCACTTCCAGGTGCG
 40 GCTCAACCAATACCTGTGTGGGTACAGCTACCATGCGAGCCGAAACCGGATGTAGCAATGTCTCACTTCC
 ATGCTCACCGACCTCTCCACATACAGCAGAAACGGCTAAGCGTAGGTGCGCAGGGGGTCTTCCCGCT
 CTTTGGCCAGCTCTTTCAGCTAGCCAGTTTGTCTGCGCTTCTTGAAGGCGACATGCACTACCCACCATGT
 CTCTCCGAGCGCTGACCTCATGAGGCCAACTCCTGTGGCGGAGGAGATGGGCGGGAACATCAACCCGC
 45 GTGAGTTCGGAGAACAAAGGTGTGATCTTGGATCTTTTCGACCCGCTTCGAGCGGAGAGGATGAGAGGG
 AAGTATTCGTTCTCGCGGAGATCTTGGCGAAATCCAAAGAGTTCCCGCAGCGATGCCATCTGGGCGGG
 CCGGATTAACAACCTTCACTGTTAGAGTCTTGGAGAGACCGGACTACGTCTCTCGGGTGTGCGAGCGG
 TGGCGTTTGCCACCTATCAAGGCCCTTCAATACACCTTCCACGAGGAAGAGGACGGTTGTCTCATGACG
 50 AGTCTCTCGTGTCTTCTGCTTACGCGAGCTCGCTACTAAGACCTTCGCGAGCTCCGAATCTGCGCCGT
 CGACACGGGACGCGGACCGCCCTTCTGACAGGCTCCGAGCAGGTGACAAAGGATCCGAGTGTGAGT
 TGTACTCTTCCATCGCCCGCTTGGAGGGGAACCGGGGACCCGATCTCAGTGGACGGGTCTTGGTCTA
 CGTGGAGCGGAGGAGCTAGTGAGGATGTCTGTCTGTCTCAATGTCTTACACATGCGAGCGGCTTGTAT
 55 CAGCCATCTGCTGCGAGGAAAGCAAGCTGCCATCAACCGTGTGAGCACTTTTGTCTGCGCCACCAT
 AACTGGTTTATGCCCAACATCTGCGAGCGCAGGCTTGGCGAGAAAGAGTTCACCTTTGACAGCATGCG
 AAGTCTCTGAGCAACACTACCGGAGCTGCTCAAGGAGTAAAGGCGTCCACATGTAAAGCTTAAAGCTTAA
 ACTCTATTCTGAGAGGAGGCTGCAAGCTGACGCGCCCACTTCCGCGCAAAATCCAGTTTGGCTATGGG
 60 GCAAGGCGCTCCGGAACCTATCCAGCAAGGCCGTTAAACCATCCACTCCGTGTGGAAGGATCTCTGG
 AAGACATGTGACCAACATGACACCACTATCCATGCGCAAAATGAGGTTTCTGTGTCTCAACACAGAGAA
 AGGAGGCGGTAAAGCAGCCCGCTTATCGTATCCCGAGATCTGGAGTCCGTGTATGCGAGAGAGATGGCC
 CTCTATGATGTGTCTTCCACCTTCTCAGGTGTGATGGGCTCTCATACGAGTTCGACATCTCTCTGTG
 GCGAGCGATCGAGTTCTGTGTAATACCTGGAAATCAAGAAAAACCCCATGGGCTTTTCTATATGACAC

gi|306286|gb|U6362.1|)HPCUNKDCS Hepatitis C virus mRNA, complete cds
TCCGACGCGCGGATTGGGGGACGACTCCACCATGATGACTCCCCCTTGGAGCAACTACTGCTCTACGCG
AAGAAAGCGCTCTAGCAATGCGGCTGTATATGATAGTGTGTGCGACAGCTCCGAGACCCCCCTCCGGGAGAGGCC
25 ATATGTGTTCTGCGAACCGTGTTAGTATACAGGAATTGTCGAGGACGACCGGGTCTCTTTCTGAGCAACAACG
GCTCAATGCTCGGTGAAATTTTGGGCGCTGCGCCGGAGACTGCTAGCCGACGTTTGTGGTGCTCGAAGAGCC
CTTGTGTGTACTGCTGATAGGTGCTGTGGAGTGTCCCGGAGGCTCTGTAGACCTGTGCACATGACGACT
GAACTCTAAACCTCAAAGAAAAACCAAACCTGTGTGCCCGCGCGAGGGCCCGACAGGATATTAAGTTTCCCGGG
GGTGGTCTGATAGCTTGGTGGATTTACTTGTGTGCCCGCGAGGGGCCCTTCCCAAGGCTCGCCGCGCCGAGG
30 GAGAAGACTTCCGACGCGTGCAGTAACCTGTGGAAAGCGCAGGCTTACCGAAGGCTCGCCGCGCCGAGG
CAGGGCCGTGGGCTGAGCCCGGTTACCTTTGGCCCTCTTATGGCTATAGAGGCTTGGGGTGGGCGAGGATGT
CTCTCTGTACCCCGCGGCTTCGGCCGCTTGTGGGCGCCACGACGCCCCGGCGTAAATCGGCTATATTGG
GTAAGGCTATCGACAACCTACATGCGGGCTTGGCCGACCTCATGAGGGTATATTCCGCTCGTCCGCGCCGCC
35 CTAGGGGGCGTTCGACGCGCCCTGGCAGCATGTTGTTGCGGGTGTGAGAGCGCGGTGAATGTATGCAACA
GGGAATCTGCCCGGTGCTCTTCTCTCTATCTTCTCTTGGCTCTGCTCTTGTGTGTAGACCCCGAGTTT
CGCGTATGAGAGTGGTAAACCGCTCGGGATGTATCATGTCGACAGCACTCCGCTCAAATCAAGCAATTGT
GTATGAGGCGCAGCGCATGATCATGACAGCACTCCGGGTGGTGTGCTCGCTCGTGGAGAGCACTCTCTCC
40 GTTGTCTGGGTGGCACTGCTCCAGCGCTCGCGCGCAGGAATGACGCGCTCCCACTACAGCAATTGCGAC
GCCATGTGCACTGTCTCTGTGGGTAGCTGCTTTCTTCTTCCGCTATGATCGTGGGGAGCTCTCGGGATAT
TGTTTCTCTTGTCTTCCAGGTGTCTACTCTTGTCTCGCCGCGCATGAGACGGTATCAGAGCTGCAACTGCG
TCAATCTCTCCCGGCGCGTACAGTACAGGTACCGCATGGCTGGGATGTATGATGAACTGGTGGCTCTCAACA
45 CAGCCCTATGTGATCGCTAGCTACTCCGGATCCCAAGATGCTGTGGTGACATGACAGGGTCCCACTGCG
GGGAATCTCGCGGGGCTTGCTACTATTTCATGCTGTGGGAGATCTGGGCTAAGGCTCTTAATGCGATGCTAT
CGTCCCTCTTTAGCTGCGTGGCGGGCTTCAGCACTCTCCAGCTATAAACAACAACGCGAGCTGGCATATCAA
46 CAGGACCGCCCTGAGCTGCAATGACTCTTCAACCTGGGTTTGTGCGCGCTTTCTTCAACAATACAGAG
TTCAACGCGTCCGGTGGCGGAGCGTGTGGCAGCTGGCGCCCATTTGATCACTTGGCGCGAGGTGGG
GTCTCCACTCATTTACATGAGGCTCATGATTTTGGATACGAGGCCCTATTTGCGGACTACGCGCTCAACG
50 GTTGGTATTTGGCCACGTTGCAGGTGTGTGGCCAGTATACCTGTTTCAACCCGAGTCTGTTTGGGTG
GGGACTACCGATGCTTGTGGTGGCCCTACATACAGATGGGGGCAATAGAGACGAGCGTGGCTGCTCTTCA
ACAACCGCGGCGCGCGAAGGACTGTTGGCTGGTGTACATGAGTAAATGACATCGGTTGACCAAGAC
ATTTGGGGGCGCCCCGTTGTAAATCTGGGGGGGTGGCAACAATCTTGTAGCTGCCGCCGAGGACTGCTTC
CGAAGAACCCCGGGCCCACTTACCAAAATCGGTTTGGGGCTTTGTAACCCAGAGTCTTTATGCGT
51 ACTACCGCTACAGGCTCTGGCACTGCGCTGCACTGCTGCACTTTCATCATCTTTAAGGTTATGAGTGTAGCT
GGGGGCGCGGAGACAGCGTTCAGCGCGCATGCTGCACTGAGTACGGGAGAGCGCTTGTGACTGTAGGAG
AGGGAATAGTGTAGAGCTTAGCCGCTGCTGTGCTGTATCAACAAGATGGGAGGTATCGCCCTGTCTCTTCA
52 CAAACCTACCGGCTGTGTCGAGCTGTTTGAATTCATCCATCAAGAACTGTGGAGACTCAATCAATCGATC

CGGTATAGGGTCGGCGGTTGTCCTCTTGCGATCAAAATGGGAGTATATTGTGCTGCTCTTCTCTTCTG
 GCGGACGCGCGCGTCTGCGCTTGCTTGTGGATGATGCTGCTGTGAGCGCAAGCCGAGGCGCGCTTAGAGA
 ACCTGGTGGTCTCAATCGAGCGTCCGTGGCCGGAGCGCATGGCACTTCTTCTCTCATTTGTGTCTTCTG
 TGCTGTCTGCTACATCAAGGGCAGGCTGGTTCCTGGAGCGGCATACGCCCTCTATGGGATATGGCCGCTG
 5 CTTCTGCTTTCTGTGGCGTTACCAACGAGGGCGTACGCCATGGACCGGAGATGGCCGATCGTGGGAG
 GCGCGGTTTTGTAGGCTGTGTACTTGTACCTGTGACCACACTATAAAGTGTCTTGTCCAGGTTCAAT
 ATGTGGTCTACAATATCTCATCACGAGAACCGAGCGCATCTGCAAGTGTGGGTCCCCCTCTCAAGGTT
 CCGGGGGGTCCGATGCCATCATCTCTCCATCGGTGGTCCACCAGAGCTATATTTAGCATCAACA
 AATATTTGCTCGCCATATTCGGCCCGCTCATGTGCTCCAGGCGGCATAACTAGATGGCGTACTTCTGT
 10 GCGCGCAAGGGCTCATTCGTGCATGCATGTTGGCGCGGAAGTCTGGGGGGTCAATACGTCCAAATG
 GTCCTCATGAAGTGGCGGCATAGCAGGTACGTACGTTTATGACCACTTACTCTCACTCGGAGATTGGG
 CTCACACGGGCTTACGAGACTTGCAGTGGCAGTAGAGCCGTTGTCTTCTCGACATGGAGACCAAGT
 CATCACTTGGGGGCGAGACACCGCGCGTGGCGGACATCATCTTGGCCCTGGCTGTCTCCGCGCAAGG
 GGGAAAGAGATACTTCTGGGACCGGCCGATAGTCTTGAAGGACAGGGGTGGCGACTCTTGTGCGCCATCA
 15 CGGCTACTTCCCAACAAACGCGAGGCTGTCTTGGTGCATCATCACTAGCCTTACAGGCGCGGACAGAA
 CCAGGTTGAGGGGAGGTTCAAGTGGTTTCCACCGCAACACAATCTTCTTGGCGACTGCATCAATGGC
 GTGTGTTGGACTGTCTCCACGCGCCGCGCTCAAAGAGCCTAGCGCGCCCAAGGGTCCAATCACCACAA
 TGTACACCAATGTAGACAGGACCTTGTGGCTGGCGCACCTCTGGGGCGGTTCCCTGACCACTG
 CACTTGGCGCTCTCGGACCTTTACTTGGTTCAGAGACATGCTGATGTCACTTCGGTGGCGCGGGGT
 20 GACGCTAGGGGAGGCTACTCCCCCGAGGCTGTCTCTACTTGAAGGGTCTCTCGGTTGTGCATGCT
 TCTGCGCTTCGGGACAGCTGTGCGCATACTTCCGGCTGCTGTATGACACCAGGGGGTGGCATGGCGGT
 GGAATCTATACCGCTTGGTCTATGGAACACTACTATGCGTCTCCGGTCTTACAGGCAATCCGCTCTCC
 CCGGCTGTACCGCAGACATTCAGTGGCCCACTTACACGCTCCACCGCGCAGCGCAAGACACTAGGG
 TGCCGCTGCATATGCGAGCCCAAGGGTACAAGGTGCTCTGCTCAAACTCCGTCCGCTCGCCCACTTGGG
 25 TTTTGGGCGTATATGTCCAAAGGCATGCTATCGACCCCAACTTAGAATCGGGGTAAAGACCATCAAC
 ACAGGTGGCCCTATACATACTCCACTATGGCAAGTCTTCTGGCGACGCTGGCGGCTCTCGGGGCGCT
 ATGACATCATATGTTGTATGATGATGCCACTCACTGACCTGACACTCAATTTATGCTATCGGCACAGTCT
 GGAACCAAGCGGACAGCGCTGAGACGCGGCTCGTGTGCTCTCCACCGCTACGCTCCGGATCGGTCACT
 30 GTGCCACACTCAATATCGAGGAGTGGCCCTGTCTAAATCTGGAGAGATCCCCCTTACGCGCAAGCCA
 TTTCCATCGAGGCTATCAAGGGGGAGGCGATCTCATTTCTGCGCAATCCAGAAAGAGGTGACGAACT
 CGCGCAAGAGTGTGAGGCTCGGACTCAATGCGGTAGCGTATTACCGGGGTCTGACGTGTCCGTATCA
 CCGACGAGCGGAGACGTTGTGTGCTGGCGACGAGCGCTTAATGACGCGGCTTACCGCGCATCTTGACT
 CATGATTCGACTGTAAATACGTGTGTACCCAGACAGTCGATTTCACTTGGACCCCACTTACCATTTGA
 35 GACGACGACGCTGCCCAAGACGCGTGTGCGCTCGCAGAGGCGAGGCGAGGCTGTAGGGGCGAGGCT
 GGCAATACACAGTTTGTGACTCTCGAGAGAACGGCCCTCGGCGATGTTCACTTCTCGGCTGTGTGAGT
 GTTATGACCGGGTTGTGCTGTGACGAACCTACGCGCGCTGAGACCTCGGTAGGTTGGGGGCTACT
 AAACACACAGGGTTGCCGCTCTGCCAGGACCATCTGGAGTCTCGGAGGTTGTCTCACAGGCTCAAC
 CACATAGATGCCACTTTTATCCAGACTAAACAGCGAGGAGAACTTCCCTACTTGTGTAGCATAC
 40 AGGCTACAGTGTGCGCCGGGCTCAAGCCCCACTTCCATCTGGGATGAAATGTGGAGGTGTCTCATACG
 GCTGAAACCTACGCTCGCGGCCAACACCCCTCTGTATAGTTAGGAGGCTCCAAATAGAGTCAAC
 CTCACACCCCAATAACAAATTCATATGACATGATGTGCGCTGACCTGGAGGTGCTACACGACACT
 GGGTGTGTGAGCGAGTCTCGCAGCTCTGGCGGCTACTGCTGACAAACGCGACGCGTGTCTATGT
 45 GGGCAGGATCATCTGTCCGGAAGCGGCTATCATCCCCGATAGGGAAGTCTCTACAGGAGTTGAC
 GAGTATGAGGAGTGTGCTCACCTCTTACTTCCGAACAGGAATGACGCTCGCGGACCAATCAAAAC
 AGAAGGCGCTCGGTTGTGCTCAAACAGCCACCAAGCAGGCGGAGGCTGTGCTGCCGTGTGTGAGTCCAA
 GTGGCGAGCCCTTGGAGACTTCTGGGCGAAGCACATGTGGAATTCATTATGGAGATACAGTACTTGGA
 GCGTTTGTCACTGTGCTGGGAACCCGCAATACGATCACCGATGGCACTTACAGACTTCACTACCCAGC
 50 GCTCACACCGAGTACTACCTCTTGTTTAACTCTTGGGGGATGGTGGTGTGCCAATCGCCCTCGGACG
 CAGCGCTGCCACGCTTCTGCTGGGCGCCGCAATGCTGGAGCGGCTGTGTGACAGTATAGGCTTGGGAAG
 GTGCTTGTGGACTTTCGCGAGGTTATGGAGCAGGCTGGCGGCGACTTGTGGCTTTTAAAGTATAGA
 CGCGGAGATGCCCTTACCGCAGGACATGCTCACTTACTCCCTGCGCATCTTCTTCCGCTGCCCTGTGT
 CGTGGGATTTGTGTGTCAGCAATCTGCTGCGCATGTGGCGCCAGGGGAGGGGCTGTGCACTGTGATG
 55 AACCGCTGATAGGCTTGCCTGCGGGGTAAACAGCTCTCCCCAGGCACTATGTGCGAGAGCGAGC
 GTGACGCGGTGTATCCAGATCTTTCAGGCTCACCATCACTCAGCTGTGTGAAGAGACTCCACCAAGT
 GATTATGAGAGTGTCTTACGCCATGCTCCAGCTGTGGCTTAAAGGAGATTTGGGACTGTGCTTGCAG
 GTGTGACTGACTTCAAGACTTGGCTCCAGTCCAAGCTCTTCCGCGGATTACCGGAGGTCCCTTTTCT
 CATCGCAACGCGGATATAGGGAGTCTGGCGGGGGCGGCATCATGACACACTGCCATCGGATCGGAGC
 ACAGATCACCGGACAGTCAAAAACGCTTTCATGAGGATCGTGGGCTTAAACCTGACGACACAGTGG

TACGGGACATTCCCCATCAACGCGTACACCAGCGGCCCTGCACACCTCCCGCGGCCAAATATTCCA
 AGGCATTGTGGAGAGTGGCCGCTGAGGAGTACGTGGAGGTCAACGCGGTGGGAGATTTTCACTACGTGAC
 GGGCATGACCACTGACAACTGAAAGTGTCCATGCCAGGTTCCGCGCCCGGAATTTCTTACGGAGGTGGAT
 5 GGAATGCGGTTGACAGAGTACGCTCCGCGCTGCAGACCTCTCTTACGGGAGGAGGTGCTGATTCAGAGTCG
 GGCTCCACCACTAGTCTGGTGGTTCACAGCTCCCATGCGAGCCCGAACCGGATGTAGACAGTGTCTCACTTC
 CATGCTCACTGACCCCTCCCAATACAGCAGAGACGGTAAAGCGTAGGCTGGCCAGGGGGTCTCCCCC
 TCCTTGGCCAGCTTACAGTAGCCAGTTGTCTGCGCTTCTTGAAGCGCAGATGCATACCCATCTAGT
 ACTCCCCGAGCGCTGACCTCATTTAGGCCAACCTCTTGTGGCGGCAAGAGATGGCGGGAACTCACCOC
 CGTGAGATCAGAGAAATAGGTGTTAATCTTGAGCTCTTTTCGACCCGCTCCGAGCGAGGATGATGAGGG
 10 GAAATATCTGTTCCGCGGAGATCTTGGCGAAATCCAGGAAATTTCCCCCGAGCGTGCCTCATGGGCGC
 CGCCGGGATTAACCCCTCCGTGCTAGAGTCTTGAAGGACCCGAGACTAGTCTCTCGTGGTGTGACAGG
 GTGCCGTTGCCGCCACCAAGGCCCTCCAATACCCACTCCACGAGGAAAGGAGGACGGTTGTCTTGACA
 GAATCCACCGTGTCTTCTGCTTGGCGGAGCTCGCTACTAAGACCTTCGGCAGCTCCGAGATCGTCGGCCA
 TCAGACAGCGGTACGGCGACCGCCCTCTTGACCAAGCCTCCGCTGACGGCGACAGAGAGTCCGAGCTTGA
 15 GTCGTTCTCTCCATGCCCCCTTGGAGGAGAGCGGGGGGACCCGATCTCAGCGACGAGATCTTGTGTC
 ACCGTGAGCGAGGAGGCTAGTGAGGACGTCGTCTGTCTTCGATGTCTTACACATGGACGCGCCCTGA
 TCCAGCCATCGCTGCGGAGGAAAGCAAGTTGCGCATCAACCCGTTGAGCAATTTCTTGTACGTACCA
 CAACATGTTCTATGTACAACAATCCCGACGCGAGCGCTGGCGCAGAAAGGTCACCTTTCAGACAGCTG
 CAAGTCTTGACGAGCACTACCGGAGCGTGTCTTAAGGAGATGAAGGCGAAGGCGTCCACAGTTAAAGGCTA
 20 AACTTCTATCTGTGAAGAGCGCTGAAACTGACGCCCACTTCGGCCAAATCCAAATTTGGCTACCG
 GGGCAGGAGCGTCCGAGCCTATCCAGCAGGGCGGTACCCACATCCGCTCCGTTGGAGGAGACCTGCTG
 GAGACAGCTGAACCAAAATAGCACATACCATATGGCAAAAATGAGGTTTCTGTGTCCAACTGCTGCC
 AAGGAGGCGCGAAGCCAGCTCGCCTTATCGTGTTCGCCAGATCTGGGAGTTCTGTGATTCGAGAGAGATGCT
 25 CTTTATGATCGTGTCTCCACCCCTTCTCAGGCGGTGATGGGCTCTCATACGATTCACGATCTCTTCCT
 AAGCAGCGGGTCGAGTTCTGTTGTAATACCTGAAATCAAGAAATGCCCATGGGCTTCTCATATGACA
 CCGCTGTCTTTGACTCAACGGTCACTGAGAATGACATCCGTTTGAGGAGTCAATTTACCAATGTGTGA
 CTTGGCCCGGAGGCCAAACTGGCCATAAAGTCGCTCAGAGAGCGCTCTATATCGGGGGTCCCTGTAGT
 AATTCTAAAAGGGCAGAACTGCGGTTACCGCCGGTGGCGGCGAGCGGCTGTGTCAGCAGCTAGCTGCGGTA
 ATACCTTCACATGTACTCTGAAAGCCACTGCGCGCTCTGAGCTGCGAAGCTCCGAGACTCAGCAGATGCT
 30 CGTGAACGAGAGCAGACTTGTCTGTTACTGTGAAGCGCGGGAACCCAGAGAGATGCGGCGAGCCTACGA
 GTCTTCACGAGGCTATGACTAGGTACTCTGCCCCCTGGGGACCCGCTCAACCCGGAATACGACTTGG
 AGTTGATTAACATCATGTTCTCCAATGTGTGCGTGCACAGATGCATCTGGTAAAGGGTGTACTACCT
 CACCCGTGACCTTACCACCCCTTGCACGCGCTGCGTGGGAGACAGCTAGACACACTCCAGTCAACTCC
 TCGTAGGCAACTCATCATGTATGCGCCACCTTATGGGCAAGGATGATTCTGATGACTCATTTCTTCT
 35 CCATCTCTTAGCTCAGGAGCAACTTGAAAAACCTAGATTGTGAGATCTACGGGCGCTGTACTCCAT
 TGAACAGCTGATCTACCTCAGATCATTTAGCGACTCCATGCTTATAGCGCATTTTCACTCTCATAGTTAC
 TCTCCAGCGAGATCAATAGGTGGCTTCACTGCTCAGAAACCTTGGGTTACCACTTTCGAGGCTTGA
 GACATCGGCGCAGAGTGTCCGCGTAAAGTACTGTCCAGGGGGGAGGGCGGCCACTTGTGGCAGATG
 CCTCTTCAACTGGGCGGTGAGGACCAAGCTCAAACCTCACTTCACTCCAGCGCGTCCCGGCTTGAAGTGT
 40 TCCGGCTGCTGTTCTGCTGTTTACAGCGGGGAGACATATATCACAGCGCTGTCTGTGCCGACCCCGCT
 GGTTCATGTTGTGCTACTCTTACTTTCGTTGGGGTAGGATCTACTGCTTCCCAACCGATGAATGGG
 GAGCTAAACATCCAGGCCAATAGGCGGTTTCTC (SEQ ID NO: 6889)

45 gi|329739|gb|U02836.1|HPCCGENOM Hepatitis C China virus complete genome
 ATTGGGGCGCACTCCACATAGATCACTCCCTGTGAGGAACACTGTCTTCAACGACGAAAGGCTCA
 GCCATGGCGTTAGTATAGTGTCTGTCGACGCTCCAGGACCCCGCTCCCGGAGAGCCATAGTGTGCTCG
 GGAACCGGTAGTACACCGGAATGCGAGGACGACCGGCTCTTCTTGAATCAACCCGCTCAATGCTCG
 GAGATTGGGCGTCCCGCGCAGACTGTCTAGCCGAGTATGTTGGGTGCGGAAAGGCGCTTGTGGTACTG
 50 CTTGATTAGGGTGTGCGAGTGCCCGGAGGCTCTGTAGACGCTGACCACTGAGCAGCAATCTTGAAC
 TCAAAGAAAAACCAACAGTAAACCAACCGCGCCGACAGGACGCTCAAGTTCCCGGGCGGTGTGTCAGATC
 GTGTGGTAGGTTTAACTGCTTTCGCGCGCAGGGGCCCAGGTTGGGTTGCGCGCGAGATGAGAGACTTCCG
 AGCGGTCGCAACTCTGTGGAGGCGGACACCTATCCCCAAGGCTTCGCGACCCGAGGGCGAGCTTGGCC
 55 TCAGCCGCGGTATCTTGGCCCTCTATGGCAATGAGGCTTGGTGGGCAAGATGCTCTGTCTCACCC
 CGCGCTCCCGGCTAGTTGGGCCCCCAGCAGACCCCGCGTAGGTGCGGTAATTTGGGTAAAGGTCACTG
 ATACCTTCACATCGGCTTGGCGGACTCATGGGTATCACTGCTGCGGGGCCCTTGGGGGGCGC

TGCCAGGGCCCTGGCACATGGTGTCCGGGTTCCTGAGGACGCGTGAACATAGCAACAGGAAATTGCCCC
 GGTGTGCTCTTTCTCTATCTCCCTTTTAGCCTTGCTATCTGTTTGACCAACCAGCTTCGCGTTACGAAG
 TCGGTAAAGTGTCCGGGATATACCATGTACGAACGACGTCTCAACTCAAGCATTTGGTATGAGGCGAGC
 GGACCTGATCATGATACACCTCGGGTGGGTGCCCTGCCTTCGGGAAGGCAACTCTCCCGGTGCTGGGTA
 5 CGGCTCACTCCCAAGCTCGCGGCAGGAACGCCACGATCCCCACTGCGACAGTACGACGCGCATGTGCAT
 TGTCTGTTGGGGCGGTGCTTCTCTCCGCCATGTACGTGGGGGATCTCTGCGGATCTGTTTTCCTTGT
 CTCTCAGCTGTCTCACTCTCTCGCCTCGCCGATGAGACAATACAGGACTGCAATGTCTCAATCTATCCC
 GGGCCAGTAACTGCTCAGGCTACCGCATGGCTTTGGGATATGATGATGAATGGTGCCTCAACAGACTCTAGTGG
 TGTCCGACTTACCTCCGATCCCTCAAGCCGTCAAGACATGTGGTGGGGGCCACTGGGGAGTCTTGGC
 10 GGGCTTTGCTACTATGCCATGGTGGGGAATTGGGCTAAGGTTTGTGATTGTGATGCTACTCTTCGCGCGC
 GTTGATGGGGAATACCTACGCGTCTGGGGGGGCGCAGGGCCGCTCCACCCTCGGGTTCACTGCCCTCTTTA
 CACCTGGGGCCTCTCAGAAATACAGCTTATAAATACCAATGGTAGCTGGCATATCAACAGGACTGCCCT
 GAATCGAATGATCTCCCTCAATACTGGGTTTCTTGCGCGCTGTCTATACACACAGGTTCAACAGCGTCC
 GGAATGGCAGAGCGATGGCCAGCTGCCGCGCCCATGTGATACATTCGATCAGGGCTGGGGGCCCATCACTT
 15 ATACTGAGCCTGATAGCTCGGACAGAGGCGCTTATTGTGTCACCTACGCGCTCGAAAGTGTGGGCATCTGT
 ACCTGCGTCGGAGGTGTGCGGTCCAGTGTATTGTTTCACCCCAAGCCCTGTGCTGTGGGGACAGCCGAT
 CGTTTCGGTGTCCCAACATATAGCTGGGGGGAGAATGAGACAGAGCTGCTGCTCTCAACAACACGCGGC
 CGCCCGAAGGCACTCGTTTGGCTGTATACATGATGAATGGCATCTGGGTTACCAACAGAGCTCGGGGGGCC
 TCCGTGTAACTCGGGGGGTGCGCAACAACACTTTGACTTGCCCCACGGATGCTTTTGGGAAGCACCCC
 20 GAGGCTACGTATACAAAGTGTGGTTCGGGGCCTTGGCTGACCTAGGTGCTTAGTTGACTACCCATACA
 GGCCTTGCCGATCACTCCCTGCACATGTCAACTTTGCCATCTTCAAGATTAGGATGTATGTGGGGGGCGTGGGA
 CACAGGCTCGATGTGCTGATGCACACTGGACTCGAGGAGAGCGCTGTAACCTTGGAGACAGGATAGATACA
 GAACTCAGCCCGCTGTCTACTGTCTACAACAGATGGCAGATATACCCCTGCGCCTTCAACACCTTACCGG
 CTCTGTCACTGTGTTAATCCACTCTCCATCAGAACTCGTGGACGTGCAATACCTGTACGTTACGTTAGGTC
 25 AGCGTTGCTCTCTTGTCAATTAAATGGGAGTAGTCTTGTGTCTTTTCTTCTACTAGCAGACGCGCGC
 GTATGTGCTGCTTGTGATGATGTCTGTATAGCCAGGCGGAGCGCGCTTAGAGAACTCGTGGTGTCC
 TCAATGCGCGTCCGTGGCGGACGCGCATGGCATCTCTCTCTTCTTGTGTCTTCTTGTGCGCGTGGTA
 CATTAAGGCGAGCTGCTGCCCGGGGAGCATACGCTTTCTACGCGGTGTGGCGGCTGTCTCTGTGCTCTG
 CTGACATACCAACCAAGCTTACGCCATGGACCGGGAGATGGCTGTGATCGTGGGAGGCGCGGTTTGTG
 30 TAGTCTGTGTTTCTGACTTTGTCAACATATCAACAAGTGTCTCTCGCTAGCTCATATGTGTGTTGCA
 ATACTTCTCAACATAGCCGAGGCGCACTTGAAGTGTGGATCCCCCTCTCAACATTCGAGGGGGCGC
 GATGCCATCATCTCTCAGCTGTGCAATCCACCCAGAGTCAATCTTTGATCATCAACAACTCTGCTGTG
 CCACGCTCGGTCCGCTCTGTGCTTCAAGCTGGCATAACTAGAGTGGCGTACTTTGTGCGGCTCATGG
 GCTGCTGCGCGCTCATGTCTATTGCGGAAAGTTGCTGGGGGTCAATTATGTCCAAATGGCTTCATGAAG
 35 GTGGGCGCATGACAGGTACGTACTGTCTAATAACCTCTTACTCGCTGCAATGTGGCCAGCGCGGGTT
 TACGAGAACTCGCGGTGCGATAGAGCCCGTCACTTCTTGTGATGAGAGCAAGATTATCACTGGGG
 GCGCAGACTCGAGCTGTGAGAGCATCTTTGGGTTTACCCGTCTCCGCGGAAGGGGAAGAGGATA
 CTCTCGGGCGCGCGATAGTCTTGAAGGCGAGGGGTGGCGACTCTTGGCGCATCAACGCTACTCTCC
 AACAGCGCGGGGCTTACTTGGTGTGATCATCACTAGCTCAACAGGCGGAGCAAGACAGGCTCGAAGG
 40 GGAGGTTCAAGTGGTCTCCACCGCAACACAACTCTTCTCGGCACTTGCATCAAGCGGTGTGTGTGACT
 GTCTATCATGCGCGCGCTCAAAAACCTTAGCGCGGCCAAAGGGCCCAATCAACCACAAATGTACACCAAT
 TAGACAGAGACTCTGCTGCGCTGGCACCGGCCCGCGGCGCTTCTTAAACCATGCACTCTGGGCGAG
 CTCGGAACCTTACTTGTGTCAGGACATGTGATGTCACTCCGGTGTGCGCGTGTGAGGCGACATGAGGGG
 AGTTTACTCTCCCCAGGCGCTGTCTCTCACTGAAGGGTGTGCGGGGCGCACTGTCTGCTCCCTTGTG
 45 GGCAGTGTGCAAGCACTTCCGGGGTGTGTGTGCAACCGGGGGGTTGCAAGGGGGGTGATTTTATAC
 CGTTGAGACCATGGAATACCATCGGCTCCCGGCTCTTCAAGACAACTGCACTCCCTCTGTCCCGTACCG
 CAGACATCTCCAGTGGCCCATCTACAGCTCCCACTGGCAGCGGCAAAAGCAACAAAGTGTGCGGGTGTGAT
 ATGCACTCCCAAGGTACAAGTACTTGTCTTGAACCGTGTGTTGCGGCGCACTTAGGTTTGGGGCGTA
 TATGCTTAAGGCACATGGTGTGCAACCCCAATTAAGAACGGGGTAAGGACCATCAACACGGGCGCCCC
 50 ATCACTACTTCACTTATGGCAAGTTCTCTGTGATGGTGGTGTGCTTGGGGGTGCTTACGATTAATAA
 TATGTGATGATGCCAATCAACTGACTCGACTACCATCTTGGGCACTCGGACCGCTCTTGGCAACAGCGGA
 GACGCTTGGAGGCGCGCTTGTGCTGCTCGCACCGCTAGCGCTCCGGATCGGTCAACGCTGCCATCCGA
 AACATCGAGGAGTGGCCCTGTCCAAATACGGAGAGTCCCTTCTATGTATGAGCCATCCCATCGAAG
 CCATCAGGGGGGGAAGGCATCTCATTTTCTGCCACTCCAAGAAAGTGTGAGAGCTTGTCTCAAGACT
 55 ATCATGCTCTGGGCTCAACGCTGTGGCTACTACCGGGGCTGTGATGTGCTGCTTACCATCACTGAGCGGA
 GACGTGTTGTCTGTGCAACGAGGCTCTTAATGACGGGCTTTACGGGCGACTTTGACTCATGATGATGCAAT
 GTAAACATGTGTTTACCAAAACATCGATTTCAGCTTGGACCCCACTTCAACATCGAGACAAACGACCT
 GCCCCAAGACCGGTGTGCGCTCGCAGCGGCGAGGTAGGACTGGCAGGGTAGGGAAGGCACTACAGG

TTTGTACTCCAGGAGAACGGCCCTCGGGCATGTTGACACTCTCAGTCTCTGTGTAGTGCTATGACGCGG
 GCTCTGCTTGTGTACGAGCTCACGCCGCTGAGACCAACGGTTAGGTTGCGGGCTTACCTTAATAACACAGG
 GTTGCCCGCTGTCCAGGACCATCTGGAGTTCTGGGAGGGCGCTTTCACAGGTCTACCCCATATAGACGGCT
 CACTTTCTGTCCAGACCAAGCAAGCAGGAGACAACTTCCCCTACCTGTGTAGCATACCAAGCTACAGTGT
 5 GTGGCAAGGCTCAGGGCCCACTCCATCTGTGGGATCAAATGTGGAAAGTGCTCACACGGCTAAAGCGCTAC
 GCTCGAGGAGCAACACCCCTGTCTGTATAGGCTAGGAGCGCTCAAATGAGGTACACCTCACACACCGG
 ATAACTAAATACATCATGACATGCTATCTGGCTGACCTGGAGGTGCTACACAGCACTTGGGTGTGGTGG
 GCGGAGTCTTTCGAGCTCTGGCCCGTATTGCTGACCAACCGGCGAGTGGTCAATTGTGGGTAGGATTGT
 CTTGTCCGGAAGTCCGCTATTGTCTCTGACAGGGAAGTTCTTACCAAGACTTCGACGAGATGGGAAG
 10 TGTGCTTACACCTTCCCTTACATCGAACCGGAATGAGCGTCCGAGCAGTTCAGCAGAAAGCGCTCTCG
 GGTGTGCTGCAACAGCCACCAAGCAAGCGGAGGCTGTCTCTCCGCTGGTGGAGTCCAAAGTGGCGAGCCCT
 CGAGACATTTTGGGAAAACACATGTGGAATTTTCATCAGCGGATACAGTACTTAGCAGGCTTATCCACT
 CTGCTGGGAACCCCAATGGCATCACTGATGGCATTCACAGCTTCTATCACCAGCCCGCTCATACCC
 AACACACCTCTCTGTTAAACATCTTGGTGGATGGGTGGCTGCCAACTCGCTCCCCCGAGCGCGCTCT
 15 GGCCTTTGTGGGCGCGGCATCTGCCGCTGCGGCTGTTGGCAGCATAGGCTTGGGAGGTTCTTGTGGAC
 ATCTCGCGGGTTATGGGCGGGGGTGGCTGGCGCATCTGTGGCTTTAAGGTCATGATGGCGAAATGC
 CCTCCACTGAGGACCTGGTTAATTACTCCCTGCCATCTCTCTCTGTGGCTTATGTCGTGGGGTCTGT
 GTGGCAGCACTACTCGCCGACAGCTGGCGCGGAGAGGGGCTGTGAGTGGATGAACCGCTGATTA
 GGGTGTGCTTGGCGGGTAACCATGTCTCCCCACGCATATGTGCTGAAAGTACGCCCGAGCGGCTGAT
 20 TTAACCGATCTCTCCAGCTTACCATCACTACGTCTGTAAGAGTCTTCCAGGTTGGATTAATGAGGA
 CTGTTCCACACCATGCTCCGCTCTGTGGCTTAAGGAGTGTGGGATTGGATATGCAACGGTGTGACCGAT
 TTCAAGACTCTGGCTCAGCTCAAGCTCTCGCGCGGTGCGCGAGTCCCTTCTCTCATGTCCAAAGCG
 GGTACAGGAGTCTCGGCGGGGGAGCGTATTATGCAAAACCACTGTCCATGTGGAGCAGAGATTACTG
 ACATGTCAAAACCGTTCATAGGAATCGTGGGCTTAAGCTTTGAGCAACACGTGCGATGGAACTAT
 25 CCCATCAACGGTACACCAACCGGCCCTGACACCTCCCGCGCGCGAATCTTCCAGGCGCTGTGGC
 GGGTGGCTCTTGAGGAGTACGTGGAGGTACGCGGCTGGGGGATTTCCACTAGTGAAGCGGCATGACAC
 CGACAGCGGTGAATGCCATGCCAAGTCCCGGCCCTGAAATTTCTTACGAGAGTGGATGGAGATCGGCTG
 CACAGGTACGCTCCGCGTCAAACTCTCTTACGAGGAGGTCTGTGTCCAGGTCCGGCTCAACCAAT
 ACCTGGTGGATCACAGCTCCCATGCGAGCCGAGCGGAGCGTACAGTGCTCACTTCCATGCTTACCGA
 30 CCCCCTCCACATACAGCAGAGACGGCCAGCGTAGGCTGGCCAGGGGGTCTCCCCCTCTCTTGGCAGC
 TCTTCAGCTAGCCAAATGTCTGCGCTTCTTTGAAGCGCATGTAATCCCATCATGACTCCCGCGAGC
 CGACCTCTATGAGGCCAATCTCTGTGGCGGAGAGATGGGCGGAACATCAACCGTGTAGGATCAGA
 AAATAGGTAGTATCTCTGACTCTTTGACCCGCTCTCGGCGGAGAGAGCAGAGGGAGTATCTCGGTT
 CGGCGGGAGTCTGCGGAAATCCAGGAAGTTCCCTCAGCGCTGCCATATGGGCAAGCCGACCACTACA
 35 ACCCTTCACTCTGTAGGCTTGGGAAGAACCGAGATTATGCTCTCCGCTGTGATACAGGTTGCCGTGTC
 GCCTACCAACGGCCCTCCAGTACACCTTCCACGAGAAAAGGACGGCTGCTCTAACAGAGTCATCCGTT
 TCTTCTGCTTGGCGAGCTCGCTACTAAGACCTTCCGCGAGCTTGAAATCTGCGGCCGTGCAACGCGCA
 CGGCGACTGCCCTCTCTGACGAGGCTTCCGCGCGCGCAGCAAAAGGATCGACGTTGATCTGATCTCTC
 CATGCCCCCCCTTGAGGAGAGCCGGGGAGCCCCGACCTACAGCAGCGGTCTGTGTTCTACCGTGTGAG
 40 GAGCGAGTAGAGACGTCTGTGCTCAATGTCTTATACATGGACAGGCGCTCTGATCAGCATGTGT
 CTGGCGAGGAGAGCAAGCTGCCATCAACCGCTGAGCAACTCTTGTGTGCTACCAACAATGGTCTA
 TGCTACACACTCCCGCAGTGCAGGCTACGCGCAGAAAGGTGCTTTTGACGAATCAAGTCTCGGAC
 GACCACTACGGGACGTCTCAAGGATGAAGGCGAAGGCGTCCACAGTTAAGGCTAACTCTATCCA
 TAGAAGAGCGCTGCAAGCTGACGCCCCACATCTAGCCAAATCCAAATTTGGCTATGGGCGCAAGAGCT
 45 CGGCAAGCTATCTCAGCAGGCGGTAAACCATCTCGCTCCGTGTGGAAGGATCTTGTGGAGAACAAAGCT
 ACACCAATCAATACCAACCATGATGCAAAAATAGAGTTTCTGCGTCCAAACAGCAAAAGGAGGCGGTA
 AGCCAGCTCGCTTATGTAATCCAGACTTGGAGTCCGCTGTGTGCGAAGAGTGGCCCTTTATGACGT
 GGTCTTCCCGCTTCTCAGCCCGTATGGGCTCTCATACGAGTTCCAGTACTCTTGGCAGCGGCT
 GAAATTCCTGCTAAATGCTGGAATCAAAGGAAAACCTATGGGCTTCTCATATGACACCCGCTGTTT
 50 ACTCAACGCTCACTCAGAAGACATCCGTGTTGAGGAGTCAATTTACCAATGTTGACTTGTGGCCCCGA
 GGGCAGACGGGCAATAAGTCTGCTCAGAGCGGCTCTATATGGGGGTCCCTGACTAATTCAAAAGGG
 CAGAACTGCGGTATGCGCGGTGCGCGCAAGTGGCGTGTGAGCAGCAGCTCGGCTAATACCCCTTACAT
 GTTACTTGAGGCGCTCTCGGCTGTGAGCTGCGAAGCTCGAGGACTGCACGATGTGTGTGAAGCGAGA
 CGACTTGTGCTTCTGTGAAGCGCGGAACTCAAGAGGATGCGGCGAGCTACGAGTCTTTCACGGAG
 55 GCTATGACTAGGTACTCTGCCCCCTGGGACCTGCGCCAAACAGAAATACGACTTGGAGCTAATAAAT
 CATCTGCTCTCAATGTGTGATGCGCCACAGTGCATCTGGCAAAAGGTTGACTACTTCAACCGTGAACC
 CATCACTCCCTCTGCGCGGGCTCGGTGGGAGACAGCTAGACACATCCAGTCAACTCTGCTGGCAGC
 ATCATCATGTATGCGCCACTCTATGGGCAAGGATGATCTGATGACTCACTCTTCTCATCTTCTGAG

10

15

20

25

30

40

45

349

CGCGCCCAACGGTCTCATTAGGCTCTGCATGCTGGTGCAGAAACAGCGGGCGGTCACTATGTGCAGATGG
 CTCTGTGTAAGACTGGGGGCACTTACTGGCACTTACATTTACAACACCTTTCCCCCACTCCAGAGCTGGGC
 TCAATGGCAGCTTGGCTGATCTAGCGGTGGCCACCGAGCCCGTCACTTCTCCGGATGGAGATCAAGACT
 ATCACTTGGGGGGCAGACACCGCGGCTGTGGAGACATCATCAACGGGCTGCCCTGTTTCTGCTCGGAGGG
 5 GGAGAGAGGTGTTGTTGGGACCAAGCCGATGCCCTGACTGACAAGGGATGGAGGCTTTAGCCCCCATCA
 AGCTTACGCCCAACAGACAGAGGCTCTCTGGGCTGTATTGTACACAGCCTTCCAGCTCGGGAGCAAAAT
 CAAGTGGAGGGGAAATCCAGATTGTGTCTACCGCAACCCAGACGTTCTGGGCCACTTGTGATCAACGGAG
 CTTGCTGGACTGTTTATCATGTGGGCGGATCGAGGACCATCGCTTCGGCGTCCGGGTCTCTGTGGTCCGAT
 GTAACCAATGTGGACAGGATTGCTGGCTGGCCAGCGCTTCCCTGACCGGCTGCT
 10 ACTGTGGGTGCTCTGGATCTGTACTTGGTCACGAGGACGCGGATGTATCCAGTGGCGGCTGCGAGGCG
 ATAAACAGGGGAAGCTTGTCTTCTCCCGGCGCCATCTCATACCTAAAGAGTCTCGGGAGGCGCTTGTCT
 CTGCCCCATGGGACATGTCCGGGCAATTTTAGGGCCGGGTGTGCACCCGTGGGTTGCAAGGCGGTG
 GACTTTGTGCCCTGTGAGTCTTACAGAGACCATGAGGTCCCCAGTGTATTGATGACAAATCCAGCCCTC
 CAACAGTCCCCAGAGTTACAGGTGGCACTATCATGACCCACTGGGAGTGGCAAGACAGAAAGT
 15 GCGCGCCCTTACGACGCTCAAGGTTACAAGTACTTGTGTGAACCCGCTCTGTTGCTGCCACCTTATAGG
 TTCGTGTCTTATATGTCAAAGGCCATGGGATTGACCCAAACGTCAGGACCGCGTGGAGACCAATTACCA
 CAGGCTCCCCCATCACTACTCTCACCTACGGGAAATTTTGGCTGATGGCGGATGCCAGGAGTGGCTA
 CGCATCATATATGTGACGAATGTCACTCAGTGGAGCGCACCTCGATTCTGGGACTTGGATGACCGCTTCTG
 GACCAAGCGGAGAGCGCGGGGTTTAGGCTCACTGTCTTGGCACCGCTCAACCATCTGGCTTGTTCACCG
 20 TGCCCAAGTCTCAACATCGAGGAAGTTGCACTGTCCGTGACGGGGAAGAACCAATTTTATGGTAAAGCCAT
 CCCCCTAAACTGACTCAAGGGGGGAGGACATCTCATTTCTGTGCTATCCAAAGAAAGTGGCAGAGCTC
 GCTGCAAGCTGTGTGGTCTGGCGCTCAACCGGTGGCCTTTTACGTTGGCTCGACGTATCTGTCACTTC
 CAACCTACAGGAGACGTGTTGTGTGTAGCGACCGACGCTTGTATGATGTGCTTCAACCGGAGTTTGGACTC
 TGTGATGAGTGCACACCTGTTGTCTGTCCAGACAGTGCAGCTTCAGCTAGACCCATATATTTCTCTATTGAG
 25 ACTTTCACCGTGCCTCCAGGACCGCGTGTCCGCTCCCAACGGAGGGTAGGACCGGTGAGGGAAAGCATG
 GTATTTACAGATATGTGTCAACCGGGAGCGCGCGCTGGCATGTTCTGACTTCGTTGCTCTTGTGAGTG
 CTATGACCGGAGTCTGCTTGTGTGACGAGCTTACACCGCGGAGCCACAGTCAAGCTACCGGCTATCGCT
 AACAACCCAGAGTTTCCCGTGTGCCAGGACCACTTGGAGTTCTGGGAGAGTGTCTTCAACCGGCTCACCC
 ACATAGTGGCCCACTTCTCTGTCCAGACGAAACAGAGTGGGGAGAACTTCCCTCACTAGTGCATACCA
 30 AGCCACCGTGTGGCTAGAGCTAGAGCTCTCCCCGTCATGGGACAAATGTGGAAGTGCCTGATACG
 CTCAGGCCCACTTCACTGGGGCTACCCCACTACTATACAGACTGGGTAGTACAGAAATGAGATCACT
 TAACCAACCCAATCACCCAATACATCATGGCTTGCAATGTCCGGGAGACTGGAGGTGCTCATAGCACGTG
 GGTGTGTGTGGCGCGCTCTAGCCGCTTTGGCCGCTTACTGCTTGTCCACAGGACGCTGTGCATAGTG
 GGCAGGATATCTAGTGGGAAGCGGCGAGTACTACCTGACAGGGAGGTCTCTTACCGAGAGTTGTATG
 35 AGATGGAGGAGTGCCCGCCACAGTCCCTACTCTGAGCAGGGGATGCAATTTGGCTGAGACATTCACAGCA
 GAAAGCTCTCGGCTGTCTCCAGACAGCATCAAAGCAAGCGAGAGCATCACTCCCACTGTCTCCGACCAAC
 TGGCAGAAACTCGAGTCTTCTGGGCTAAGCACTGTGGAACCTTCGTTAGCGGATACATACTTCTGCGG
 GCCTGTCAAGCTGTCCCGGGAACCCCGCTATAGGTCGCTGATGTCTGTTTACCGGCGCGGTGACGAGTCC
 ACTAAGCACCCACGAAACCCCTCTCTTTAACAATCTTAGGGGGTGGGTGGCGCGCCAGCATTTGCTTCCCA
 40 GCTGCGGCTACTGCTTTTCTCGTGTCTGTGATTAATCTGGGCGCGGTGTTGGCAGTGTGGGCTTAGGAAAG
 TCTTAGTGGACATTAATGTGCTACGGGCGTGTGTGGCGGGGCGCTCGTGGCTTTCAAAATCACTGAGT
 CGGGAGACCCCAACACCGAGGATCTAGTCAACCTTCTGCTCCCATCTGTGCGACAGGAGCTCGATT
 GTGGCGTGTGTGGCGCAATATCACTCGCGGACGTCGTGGGCTTGGCGGGGCGCGGTGCACTGGATGA
 45 ACCGGCTGATAGGCTTGTCTTCTGGGTTAACCACTGTCTCCCTACACACTAGTGTGGGACGAGACGCG
 GTGCGCTGTGTGACACAAATCTCACAGGCTCACTGTCTTCTGAGTAAAGATGTCTGGGACTGAGTACGAG
 ATAAGCTCGAATGATCTCGCCCGGTGTCTAGTGTCTTGGCTTAAAGATGTCTGGGACTGAGTACGAG
 TGCTGAGGACCTTCAAGAAATGGCTGAGGSCAAACTTGTACCAACAATGCGCGGAGTCCCATCTGTATC
 CTGCAACAGCGGTTACGTTGGGCTTGGCGGGGAGGACATGTGCACTGTCTGGCTGAGTGTGGGCGG
 50 AATATAACTGACATGTCAAGAACGTTTCGATGAGAAATGTGCGGCTTACAGCTTGCAGCACACTTGGC
 GTGGGTGTTTCCCAATTAAAGCTTACACTACAGGCCGTGACGCGCTTCCAGCGCGGAGATCACTAGT
 CGGCTATGGAGGGTGTCTGACAGAGGATATGTGGAGTTAAGCGCGCTGGGGGACTTCCATTAGCTCACG
 GGGGTGACCACTGATAAACTCAAGTGTCCATGCGAGGTCCCTCACCGAGTCTTCTTACAGAGGTGGAG
 GGGTGGCTCTGATAGTACGCCCCCGCTGCAAAACCCCTGCTGGAGAGAGGTTGACGTTAGACTGG
 55 GCTCAATGAATATCTGTGTGGGTCCAGTGTGCGCTGCGAGCGCGAGCCAGAGTGTAGTGTACTGACATCA
 ATGCTACAGACCCCTCCCACTACTGACAGAGACGCGACGCGTATGCTTGGGACTGAGGCGGTCTCCCCCT
 CCGTGGCGAGCTTCTCGCGAGCGAGCTGTGCGCGCGCTCACTGAAGGCAACATCGACCACTACCAACGA
 CTCTCCAGACGCTGACCTCTAGAAGCCAACTCTGTGGAGACAGGAGATGGGGGGAACATCACTAGG
 GTGGAGTCGGAAGAACAGATTGTCTTCTGGATTCTTTCGACCGCTCGTAGCGAGAGGATATCGGG

AGATCTCTATTCCAGCTGAGATTCTGCGGAAGTTCAAGCAGTTTCTCCCGCTATGCCCATATGGGCAGC
 GCCAGATTATAATCTCTCCCTTGTGGAAACCGTGGGAAGCGCCCGGACTATGAGCCACCTTATGCTCAGCGG
 TGCCCCCTACCCACTCCCAAGCCAACTCCGGTGCCTCCACCCCGGAGAAAGAGGACGGTGGTGTGAGC
 AGTCTACAGTATCACTGCTCTGGCTGAGCTTGCCACTAAGACCTTCGGCGAGCTTCTCAACCTCAGGCGT
 5 GACAAGTGGTGAAGCGACTGAATCGTCCCGGCGCCTCTCTGGCGGTGAGCTGGACTCCGAAGCTGAA
 TCTTACTCTCCCAATGCCCTCTCGAGGGGGAGCGGGGGACCCGAGTCTCAGCGACGGGTCTTGGTCTA
 CCGTAGCGCAGTGTAGTGGTGGCAGGGAAGCGTTGTGTGCTGCTCGATGTCTTACTCTGGGACGGGCGCTTT
 AATTCAGCCCTGTGCTCTCAGAGAAGCCAAAGTCCCTATCAACGCTATTGAGCAACTCGCTGTCTCGCCAC
 CACAACCTTGGTGTATTCCACCACTCTCCGAGCGCTGGCCAGAGACAGAAAAAGTTACATTTTGAAGAG
 10 TGCAGTCTGTGTAAGACCAATTACCGGAGCTGTCTCAAGGAGGCTTAAGGCCAAGGACTCAAGGTGAAGCG
 TAGACTGCTATCCGTTGAGGAAGCGTGTAGCTGACGCCCCACACATCCGCCAGATCAAAATTTGGCTAT
 GGGCGGAAGGATGTCCGAAGCCAATCCAGTAAGGCTATACGCCACATCAACTCCGTGTGGCAGGACCTTC
 TGGAGGACAAATACAAACCCATAGACACTACCATATGGCAAGAATGAGTCTTCTGTGTGAAGCCCGA
 AAGGGGGGCGCGAAGCCCTCGTCTTATCGTGTACCCGACCTGGGAGTGGCGGTATGCGAGAAAGAGG
 15 GCTTTGTATGACGTAGTCAACAGCTCCCATTTGCCGTGTAGGGAGCTCTCTCAGGGTTCCAGTACTCAC
 CAGCGACCGGGTGCATCTTCTGCTTAAAGCGTGGAAATCTAAGAAAGTCCCATGGGGTTTCTCATGTA
 CACCCTGTCTTTGACTCAACAGTCACTGAGGCTGTATCCGTAAGGAGGAAGCTCTACCAATCTTGT
 GACTTGGCCCTGAGCTCGCATAGCCATAAGGTCCCTCAGAGAGGCTTTACATCGGAGGCTCAACTCA
 CCAATTTTAAGGGACAAAATCTCGGCTATACGGGATGCCGCGCAAGCGGCTGCTGACCACTAGCTGGCG
 20 TAACACCATAACTGCTTCTCTCAAAGCCAGTGCAGCCTGTGAGCTGCGAAGCTCCAGGACTGCACATG
 CTGCTGTGGCGGACGACCTCGTCTGTATCTGTGAGAGCGCCGTGTCCAGGAGGACGCTGCGAGCCCTGA
 GAGCCTTACAGGAGGCTATGACAGGACTTCCGCCCCCGGGAGACCCGCTCAACAGAAATACGAGCT
 TGGACCTTTAATACATCTGCTCTCTCAATGTGTGGTGCGGCGCGAGCGCTGGCAAGGCTTATTAT
 CTGACCCGTGACCTTGAGACTCCCCCTGCGCGTGCCTGTGGGAGACAGGAAGACACTCCAGTGAAT
 25 CCGGACCGGCAACATTTTATCATGTTTGCCTCCACTCTGTGGGTACAGAGCTCAGCGGCTTTTCCCTCAAGTT
 TCTTCATATCTATGACTCAGGAGCACTTTGGAAGGCTCTAGATTGTGAAATCTATGGAAGCCGTACACTCC
 TCTCAACCGTTGAGACTTACTGAAATCATCCAAAGACTCCAGAGCTCAGCGGCTTTTCCCTCCACAGTT
 ACTCTCCAGGTGAATCAATAGGGTGGCTGCATGCTCAGGAAGCTTGGGGTTCCGCCCTTTCGAGCTTG
 GAGACACCGGGCCGAGCGGTTCGCGCCCACTCTATCCAGGGGGGAAAGCGGCTATATGCGGTAAAG
 30 TACCTCTTCAACTGGGCGGTGAAACCAAATCAAATCACTTCACTTACCCTCCATGTCTCAGTTGGAAT
 TGTCCAACTGCTTACCGGGGTTACAGCGGGGAGACATTTATCAGAGCTGTCTCATGCCCGCGCCG
 TTTGTCTCTGTGCTACTCTCTACTTCTCAGTAGGGGTAGGATCTATCTCTCTCCCAACCGATAGAAG
 GNTGGCGAACACTCCGGTCTTTAGGCCCTATTAAACACTCCAGGCTTTAGGCCCGGT
 (SEQ ID NO: 6691)

gi|23510419|ref|NM_000043.3| Homo sapiens tumor necrosis factor receptor
 superfamily, member 6 (TNFRSF6), transcript variant 1, mRNA
 CTACCCCGCGCGAGGCCAAGTTGCTGAATCAATGGAGCCCTCCCCAACCCGGGCGTCCCCAGCGAGGC
 40 TTCTCTCCCATCTCTCTGACCAACCGGCGCTTTTCTGTGAGCTCGTCTCTGATCTCCGCGCAAGAGTGACACA
 CAGGTGTTCAAAGAAGCTTCTGGGGAGTGAAGGGAAGCGGTTTACAGTGTACTTGGCTGGAGGCTCAGGGG
 CGGGCACTGACCGAGAACACACCTTGAGCGCAGCCCTGGCTGGCTCCCAAGCGGAGCTACCTCTCTCCCGCG
 GGTGTGTGAGCCGCTCAGTACGAGGTTGGGGAAGCTCTTTTCACTCCGAGGATTTGCTCAACAAACATGCT
 TGGGCATCTGGACCTCTCTACCTCTGGTCTTCTACGCTGTGTGCTAGATTATCGTCCAAAGGTGTAAATGC
 45 CCAAGTGACTGACATCAACTCCAAGGATTGGAATTGAGGAAGACTGTACTACAGTTGAGATCAGAGAAC
 TTGAGAGGCTTGACATCATGATGGCCAAATCTGCCTAAGCCCTGTCTCCAGGTGAAGGAAAGCTAGGG
 ACTGCAAGTCAATGGGATGAACAGAGACTGCGTGCCTGCCAAGAGGGAGGAGGTACACAGACAAAGC
 CCAATTTTCTCCAAATGACAGAAGATGTAGATTGTGTGATGAAGGACATGGCTTAGAGTGGAAATAAAC
 TGCACCCGAGCCAGAAATACCAAGTGCAGATGTAAACCAAATCTTTTGTAACTCTACTGTATGTGAAC
 ACTGTGACCTTGCACCAAATGTGAATCGGAATCATCAAGGAATGCACACTCAACGACCAACCAAGAGTG
 50 CAAAGAGGAAGGATCCAGATCTAACTTGGGGTGGCTTTGTCTTCTCTTTTGGCAAATTCACATTAATGTT
 TGGGTGAAGAGAAAGGAAGTACAGAAAAATCGAGAAGCAGAGAAAGGAAACAGAGTTCTCATGAAAT
 TCTCAACCTTTAAATCTGTAAACAGTGGCAATAAATTTATCTGATGTTGACTTGAGTAAATATATACCAAT
 TATTGTCTGAGTCAATGACACTAAGTCAAGTTAAAGGCTTTGTTCGAAAGAAATGGTGTCAATGAAGCCAAA
 ATAGATGAGTCAAGAATGCAATGTCCAAAGACAGCAGAACAGAAAGTTCAACTGCTCTGTAATTTGGC
 55 ATCAACTCTATGGAAGAAAGAGCGTATGACACATGATTAAGATCTCAAAAAGGCAATCTTTGTATC
 TCTTGAGCAAAAATCAGACTATCATCTCTCAAGGACATTACTAGTGTCTCAAAAATTCAAATCTCAGA
 AATGAATCCAAAGCTTGTCTAGAGTGAAGAACAAATTCAGTTCTGAGTATATGCAATTAGTGTTT

5 GAAAAGATTCTTAAAGTAGCTGGCTGTAATACTGCTTGGTTTTTACTGGGTACATTTTATCATTTATAG
 CGCTGAGAGAGCCAAATATTTGTAGATTTTAAATATCATCATGATTCCTGCCCTCAAGGATGTTTAAATCT
 AGTTTGGGAAAACAACTTTCATCAAGAGTAAATGCAGTGGCATGCTAAGTACCCAAATAGGAGTGATGCA
 GAGGATGAAGAGTTAGATTTATGCTCTGGCCATCTAACATATGATTCCTGTAGTATGAATGTAATCAGTGTA
 10 TGTTAGTCAAAATGCTATCCACAGGCTAACCCCATCTATGAATCAATAGAAGAGCTATGACCTTTTG
 CTGAATAATCAGTTACTGAACAGGAGGAGCCACTTTGCCCTAAATTAACCTCTGATTAATTTAGAGATTTT
 ACCATATTTCTAAACTTTGTTTAACTCTGAGAAGATCATATTTATGTAAGATTAATGTAATTTGAGTGC
 AGAATTTAAATAGGGCTCTACCTCAAGACCTTTGACAGTTTATTTGGTGTATATTTATCAATATTTCA
 ATTGTGAATTCACATAGAAAACATTAATATTAATGTTTGAATTTATATGTTATGCTATTTACTGCT
 15 CTCAAAATCTACCTTCTCTTCTCAGGCATCAAAAGCATTTAGACAGGAGATTTACTAGAGCTTTTGC
 CACCTCTCCATTTTGGCTTGGTGCTCATCTTAATGGCCTAATGCACCCCAACATGGAATATCACCA
 AAAAAATCTTAATAGTCCACAAAAGGCAAGACTGCCCTTAGAAATTTAGCCTGGTTTGGAGATACTAA
 CTGCTCTCAGAGAAGAGTAGCTTTTGTGACATGTCAATGAACCATGTTTGCATCAAGAGATGATAAAATAGA
 TTTCTTATTTTTCGCCACCCCAAAAATGTTCAATAATGTCCCATGTAAACCTGCTACAAAATGCGAGCT
 20 TATACATAGCAATGGTAAAAATCATCATCTGGATTTAGGAATTTGCTCTGTCTATACCCCAAGTTTCTAAG
 ATTTAAGATTTCTCTTACTACTATCTACGTTTAAATATCTTTGAAAGTTTGTATTAATGTGAATTTTA
 AGAAATAATATTTATTTCTGTAAATGTAACTGTGAAGATAGTTATATAACTGAGACGATACCTGGAA
 CCACCTAAAGAACTCCATTTAGGAGGATTTTTCGCCCTTGTGTTTGAAGTATATAAATATAGGTAA
 AAGTACGTATTAATAATATGTTTTTGTGTAACAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA
 25 AAAAAAAAAAAAAAAAAAAAAA (SEQ ID NO: 6692)

gi|35910[emb|X12387.1|HSRCYP3 Human mRNA for cytochrome P-450 (cyp3 locus)
 25 GAATTCCTCAAGAGCAACACAGAGCTGAAGGAGAGATTAAGTAAAGAAATAGTGA
 TGGCTCTCATCTCCAGACTTGGCCATGGAACCTGGCTTCTCTGGCTGTAGCTGTGGTCTCTCTATCT
 ATATGGAACCATTCACATGGACTTTTAAAGAGCTTGAATTCAGGGCCACACCTCTGGCTTTTGTG
 GGAAATATTTTGTCCCTACCATAGGGCTTTGTATGTTTGACATGGAATGTCATAAAAGATGAGAAAG
 TGTGGGGCTTTATGATGGTCAACAGCTGTGTGCGCTATACAGATCTTGACATGATCAAAACAGTGCT
 30 AGTGAAGAATGTTATTTCTGTTTCAAAACCGAGGCTTTTGGTCCAGTGGATTTATGAAAAGTGCC
 ATCTCTATAGCTGAGGATGAAGAATGGAAGATTACATCATTTGCTGTCTCAACCTTCCACAGTGGA
 AACTCAAGAGAGATGGTCCCTATCATTTGCCAGTATGGAGATGTGTTGGTGAAGAACTTGAGGCGGGAAGC
 AGAGACAGGCAAGCTGTGCACTTGAAGACGCTTTTGGGGCTCAGCATGGATGTATCACTAGCACAT
 TCATTGGAATGAAATCGACTCTCTCAACAATCCACAAGACCCCTTTGTGGAAGAACCAAGAGCTTT
 35 TAAGATTTGATTTTGTGATCCATTTCTCTCAATAACAGTCTTTCATTTCTCATCCCAATCTTGA
 AGTATTAATATCTGTGTGTTTCCAAGAGAAGTTACAAATTTTAAAGAAATCTGTAAAGAGGATGAAA
 GAAAGTCCCTCTGAAGATACACAAAGACCGAGTGGATTTCTCTCAGCTGATGATTTGATCTCAGAAT
 40 CAAGAAGAACTGAGTCCCAAAAGCTCTGTCCGATCTGGAGCTGTGGCCCAATCAATATCTTTATTT
 TGCTGGCTATGAACCAAGAGCAGTGTCTCTCTCTCATTTATGATGAATGGCCACTCACCCTGATGCT
 CAGCAGAAATCGCAGGAGAAATGATGCAAGTTTACCATAAGGCACCCACCTCTGATGATCTGTG
 TACAGATGAGATATCTTGACATGGTGGTGAATGAACGCTCAGATTAATCCCAATTTGCTATGAGACTTGA
 GAGGCTCTGCAAAAGATGTTGAGATCAATGGATGTTCTTCCAAAGGGTGGTGGTGAATGATTTCA
 45 AGCTATGCTCTTACCGTGACCCAAAGTACTGACAGAGCTGAGAAGTTCTCCTGAGAAGATTCAGCA
 AGAAGAAACAAGCAACATAGATCTTACATATACACACCTTTTGAAGTGGACCCAGAAATCGATATGG
 CATGAGGTTTGTCTCTCATGAACATGAAATTTGCTCTAACTAGAGTCTTCTGAAAGACTTCTGCTTTGCTCT
 50 TGTAAAGAAACAGATCCCCCTGAAATTAAGCTTAGAGGAGTCTTCTCAACAGAAAGACCCGTTGTTCT
 TAAAGTTTGAGTCAAGGATGACACCGTAAGTGGAGCTGAAATTTCTTAAGACTTCTGCTTTGCTCTT
 CAAGAAATCTGTGCTGAGAACACAGAGACTTCAAAATTAATTTGTGAATGAACTCTGAAATGAAGAT
 GGCCTCTACCTATGAGCTGATATAAATACCGGGATTTCTGTATGCTGATGCTGCTCATTTGTCTGT
 TAGAGTTGTATCTTGGGAATAAAGGAGGTGACCAATCAGTGTGAGGAGGTAGATTTGGCTCCTCTG
 55 CTCTCAACGGATCAATTTCCACACCCCAATTAGCACAATAACTCTCTGATGCTCTGATAGAGAACT
 CAACATTTCTCAATTAATTTCCCTCCACAAATATTAATGAATAAGAAATTTTGTATGGCTCTCAACAT
 GACATTTATATCAGATGTTTCTCTGGAGTATCTATAGTTTATGTTAAATCAATTAAGACCACTTTAC
 AAAATATATCAGATGCTTCTCTGCACTTAAGGAGAACTCTAAGAACTGAATGAGAACCAACAGTAA
 ATATTTTGGTCATTTGAATCAGTTTGGCGTGGGGCTTTGTGAGAAATATTCATGTTTAAATATGCTCTT
 TGTATCAGGCATGTGCTCAGCCTGTAACTTAGCAGTTTGGGAGGCTGAGCCGGTGGATGCTCTGAG

- 5 **g**TCAGGAGTTTCAAGACAAGCCTGGCCTACATGGTGAACCCCACTCTCTACTAAAAATACAAATTAGCT
 AGGCATGGTGGACTCGCCTGTAATCTCACTACACAGGAGGCTGAGGCAGGAGAATCACTTGAACCTGGGA
 GCGGGATGTTGAAGTGAAGTGGAGATTGCACCACTGCACCTCAGTCTGGGTGAGAGTGAGACTCACTTA
 AAAAAATATGCTCTTTGAAGCACGTACATTTTGAACAAAGAACTGAAGCTCTATTATATTATTAGTT
 TTGATTTAATGTTTTCAGCCCATCTCTTTTCATATTTCTGGGAGACAGAAAAACATGTTTCCCTACACCTC
 TTGCTTCCATCTTCAACACCCCACTGTCTCGATGCAATGAACACTTAATAAAAAACAGTGGATTGGTCAA
 AAAAAAAAAAAAAAAAAAAAAAAGATTTC (SEQ ID NO:6693)
- 10 **gi|339549|gb|M19154.1|HUMTGFB2A Human transforming growth factor-beta-2 mRNA, complete cds**
 GCCCCTCCCGTCAGTTCGCCAGCTGCCAGCCCCGGGACCTTTTTCATCTCTTCCCTTTTGGCCGGAGGAGC
 CGAGTTTCAGATTCCGCCACTCGCACCCGAGACTGCACACTGAATCCACTTCTCTCTCTTTAAATTTATT
 TCTACTTAATAGCCACTCGTCTCTTTTTCCTCATCTCATGTGCTCCAGAAATTTTTCCTCTTACTCG
 15 **CC**AAAGTCAGGGTTTCCCTCTGCCCGTCCCGATTAAATATTTCCACTTTTGAACACTTGGCCCTTTCTTT
 TTAAAGGAATTCAGACAGGATACGTTTTCTGTTGGGCATTGACTAGATTGTTTGCAGAAATTTTCGATCT
 AAAAAACAACAACAACAAAAACAAACAACCTCTCTTGATCTATACCTTTGAGAATTTGTTGATTTCTTTT
 TTTATCTGACTTTTAAAAACAACCTTTTTCACCTTTTAAAAAAATGCACTACTGTGTGCTGAGCCG
 TTTCTGATCTCGCATCTGGTCAAGGTGCGCGCTCAGCCTGTCTACCTGACGACACTCGATATGGACAG
 20 **TT**CATGCGCAGAGGATCGAGCGGATCCGCGGGCAGATCCTGAGCAAGCTCAGAGCTCAACAGTCCCCAG
 AAGACTATCTTGAGCCGAGGAGGAGTCCCCCGGAGGTGATTTCATCTCAACAGCACCAGGAGCTTGCT
 CCAGGAGAAGCGAGCGGAGCGGCGCGCTGCGAGCGGAGAGGAGCGACGAAGAGTACTACGCCCAAG
 GAGGTTTACAAATAGACATGCCGCCCTTCTCCCTCCGAACTGTCTGCCAGTGTGTTACAAACACCTT
 CTGCGTCAGTGGGCACTTTGTGCTCCAGACAGTCCAGGTGCTCTGGGTACTCTTGATGCCATCCCGCC
 25 **CA**CTTTCTACAGACCTTACTCTCAGAAATTTGTCGATTTGACGTCTCAGCAATGGAGAGAATGCTTCCAAT
 TTGTTGAAGCAGAGTTTCAGAGTCTTTCTGTTTCAGAAACCCAAAAGCCAGAGTGCTTGAACAACGGATTG
 AGCTATATCAGATTCTCAAGTCCAAAGATTAACTATCTCCAAACCGGCTACATCGACAGCAAAATTTG
 GAAAAACAAGACAGAGGCGAATGGCTCTCTTCGATGTAACTGATGCTGTCATGATATGGCTTCAACAT
 AAGACAGGAACCTGGGATTTAAATAAGCTTACACTGTCCCTGTGCACTTTTGTACCATCTAATAATT
 30 **AC**ATCATCCCAATAAAGTGAAGAACTAGAAGCAAGATTGCAAGTATTGATGGCAGCTCCACATATAC
 CAGTGGTGATCAGAAAACTATAAGTCCACTAGGAAAAAAAACAGTGGGAAAGCCCACTCTCTCTGCTA
 ATGTTATTGCCCTCCTACAGACTTGAGTCAACAACAGACCAACCGGCGGAAGAAGCGTGCTTTGGATGCGG
 CCTATTGCTTTAGAAATGTGCAAGATAATTGCTGCCCTACGTCCACTTACATTGATTTCAAGAGGGATCT
 AGGGTGGAAATGGATACACGAACCCAAAGGGTACAATGCCCACTTCTGTGCTGGAGCATGCCCGTATTATTA
 35 **TG**AGTTTCAAGACTCAGCACAGCAGGCTCTGAGCTTATATAATACCATAAATCCAGAAGCACTGCTT
 CTCTCTGTCTGCTGTCCCAAGATTTAGAACTCTAACCATTTCTCTACTCATTTGGCAGAAACACCAAGAT
 TGAACAGCTTTCTTAATATGATTGTAAAGTCTTGCAATGCAGCTAAAAATCTTGGAAGAGGTGGCAAGACC
 AAAATGACAATGATGATGATTAATGATGATGACGACGAAACAGATGATCTTGTAAACAAGAAAAACATAAGA
 GAGCCTTTGGTTTCATCAAGTGTAAAAAATTTTGGAAAAAAGTTGAAGGCCCTTATTTCACTTTTCACTACTTT
 40 **GT**TCGTTTGTAAAACTGGCATCTGCACAAAAAAGTTGAAGGCCCTTATTTCACTTTTCACTACTTTT
 GTAAGTGAGAGAGACAAGAGCAATTTTAAAAAGAAAAAATAAACACTTGGAGAAATTTATTATGTG
 TAATTTATGTGAACAACGACAAACAACAACAACAAACAGGAAATCCCATTAAGTGGAGTTGCTGT
 ACGTACCGTCTCATCCCGCGCTCACTGTATTTTCTGTATGTCTATGCAATAGGACACCTTCCCATCT
 TTAATCTTAGAGTTTAAACAGTGAGTTATTATTGTGTGTACTATATAATGAACGTTTTCATGCGCTTGGAA
 45 **AA**ATAAAAACAGGTGTATAAGTGGAGACCAATACTTTGCCAGAACTCATGATGGCTTAAGGAACTTG
 AACTCAAACGAGCCAGAAAAAAGAGGTATATAATGGGATGAAACCAAGTGAAGTTATATATGACC
 GAGAAAGTCTGCATTAAAGATAAAGACCTTGAACACATGTTATGTATCAGTGCCTCAAGGAAGCTCTT
 GTAAGTCCAAAACTAAAAAGACTGTTAATAAAGAACTTTTCAGTCAG (SEQ ID NO:6694)
- 50 **gi|186624|gb|J04111.1|HUMJUNA Human c-jun proto oncogene (JUN), complete cds, clone hCJ-1**
 CCCGGGAGGGGACCCGGGAACAGAGGGCCGAGAGGCGTGCAGGCGGGGAGGGTAGGAGAAAGAGGG
 CCCGACTGTAGGAGGCGACGCGAGCATTTACCTCATCCCTGTAGCGCTCCGCGGGCCAGAGAAGAATCTTC

TAGGGTGGAGTCTCCATGTTGA CGGGCGGGCCCGCCCCCTTGAGAGCGACGCGAGCCAATGGGAAGCCCT
 TGGGGTGACATCATGGGCTATTTTAGGGGTGTGACTGGTAGCAGATAAGTGTGTAGCTCGGGCTGGATAA
 GGCTCTCAGATTGTCACTTGAGTTGGCTGAAGCAGCGAGGCGGGAGTGGAGGTGCGCGAGTGTCAAGCAGAC
 AGACAGACACAGCCAGCGACCGAGGTGCGGAGTATAGTCCGAAGCTCAAATCTTATTTTCTTTTCACTT
 5 CTCTCTTAAGTCCCGCAGACTGACGCTGTGGCTGTGGCTCCCGGGCTGGTGGTTCCGGAGTGTCCAGAGAGCTTGT
 TCTCAGCCGGGCGCGGGAGGAGAGCCCTGCTGCCAGCGCTGTGTGAACGCGGCGAAGAGCAGCGGTAC
 CCCACGCGCCGCGCGGGGACGTCGCGCAGCGCTGACAGCAGCAGAACTTTCCGCGCGGGAGAGACCG
 GAGACAGTGTGACAGTGTCCCGGAGCGAACTTTTGCAAGCCTTTTCTGCGTCTTAGGCTTCTCCACGCGG
 TAAAGACCAAGAGGCGCGGAGAGCAGCAGCAGCAAGAGAGGAGGACGTGCGCTCAGCTTCTGCTCGACCGGT
 10 TGTGTAACTTTGGCGGCGCGAGCGCGCTGCGCGGCGCCCCCTCCCCCTAGCAGCGAGGAGGAGGACA
 AGTCGTGCGAGTCCGGGCGGCGCAAGACCCTGCGCGCGCGCGGCTCAGCGAGGTCCGCACTGATCTGCTCC
 GCGGGGAGAGCCGCTGCTCTGGGAAGTGTGTTGCGCTTGGGACTCCGAGGAACCGCTGCGCCCGAAGACG
 GCTCAGTGTGAGTCCGCGCACTTTTCAAAGCGGGTAGCGCGCGGAGTCCAGCAAGTAAAGAGTGGCGGAGG
 CATCTTAAATTAACCTGCGCTCCCTGGAGCGAGCTGTGTGAGGAGGCGCAGCGGGGACGACGACGCGG
 15 GTGCGTGTGCTCTTAGAGAACTTTCCCTGTCAAAGGCTCCGGGGGCGCGGGTGTCCCCCGCTTGGCAG
 AGCCCTGTGCGGCGCCGAACTTTGTGCGCGCAGCGCAAACTAACTCACTGTAAGTACGAGACTGTCTT
 ATGCTGTCAAAGATGGAAACGACCTTCTATGACGATGCCCTCAACGCTCTGTTCTCTCGCTCGAGAGCG
 GACCTTATGGTACAGTAAACCCCAAGATCTGAAACAGAGCATGACCTGAACTCACTACCAACCAACGCGACCC
 20 GTGGCGTCCGCGAGCTGGAGCGGCTGTAATACCTCAGTCCAGCAACCGGAGGCTGCGCGAGCGGCTGCTCAAG
 CCACCCAGTTCCTGTGCGCCCAAGAACGTGACAGATGAGCAGAGCGGGGTTCCGCGAGGCTTCTGTCGCGC
 CTTGCGCGAAGTGCACAGCGAGAACACGCTGCGCAGCGTCACTGCTGCGCGCGCAGCGCGCTCAACGCGGCA
 GGCATGTGGCTCCCGCGGTAGCTCTGCTGTGCGAGGCGCAGCGCGCGCGCTTCAAGCGCAGCTG
 25 CACGCGAGCGCGCGTCTACGCAAACTCAGCAACTTCAACCCAGGCGCGCTGAGCAGCGCGCGCGCGGCG
 GCGCTCTTACGCGCGCGCGCGGCTGCGCTTTCCCGCGCAACCCAGCAGCAGCAGCAGCGCGCGCAGCAC
 CTGCCCCACAGATGCGCCGTGACAGCAGCGCGCGCTGCGAGGCTTGAAGGAGGAGCTCAGACAGTGGCGG
 AGATGTCCGCGCAGACACCGCGCTGTCCCCATGACATGGAATCCAGAGCGGATCAAGCGCGAGGAG
 GAGCGCTAGAGGAGCGCATGCTGCTCCCAAGTGCAGAAAAGGAAGCTGGAAGAGATTCGCGCGGCTG
 30 GAGGAAAAGTGTAAACCTTGAAGCTCAGAACTCGGAGTGTGCGCTCAACGCGCAACATCTCAGGGAAC
 AGGTGGCAGCTTTAAACAGAAAGTCTGAACCACTTAACTAGTGGGTGCCAACTCATGTCTAAGCAGGAC
 GTTCCAAACATTTTGAAGAGAGACCGTCCGGGGCTGAGGGGCAACGAAGAAAATAAACACAGAGAGA
 CAGACTTGAGAACTTGAAGAAGTTCGCGCGGAGGAAAAGAGTGTCCGAGAACTTAAAGCCAGAGGTAT
 CCAAGTTGAGACTGGGTTCGCTGACGCGCGCCCGCTGCGGCTGCGCGCTTTCGCGGACCGGCTGTCCCGCGCG
 35 CCTTGGCGTGGAGCCAGGAGCGCGCGCTTGGCGGCTGCGCGCTTTCGCGGACCGGCTGTCCCGCGCGG
 AACGGAACGTTGAGACTTTCGTTAAACATTGACCAAGACTGCACTGACCTAAGCTTCACTTCACTGAT
 TTAAGGGGGGAGGGGAGGGGTTACAAACTGCAATGAGACTGTAGATTGCTTCTGTAGTACTCTT
 AGAACACAAAGCGGGGGAGGGTTGGGAGGGGCGCGAGGAGGAGGTTTGTGAGAGCGAGGCTGAGCT
 40 ACAGATGAATCTTCTTGGCTGCTTTCGTTAACTGTGTATGATCATATATATATTTTAAATTTGATTA
 AAGCTGATTAATGTCAATAACAGCTTCACTGCTTGTGAAGTATTTCTTGTGTGTGTGTGGGTATCC
 TGCCGAGTGTGTGTGTAAATAAGAGATTGGAGCACTCTGAGTTTACATTTGTAATAAAGATATATAAT
 TTTTATTTGTGTGTGTGTGTGAAATTCAGAAAGGATATTTAAGAAATAACAATAAATTTGGAAAGTA
 CTCCTTACCTTCTTCTGCACTCATCTGTAGATCTTAGTCTATCTAGGTGAGGTTGAAGAGTTAAGAA
 TGCTGTGATAAATCACTCTAGTGTCTTCTATATTAAGCAGTAAAACTGTCTCTTATAGACTTAGAA
 45 ATAAAGTCTGATGTACTGATGTCTATGTCAAGCTTCTATCTCACGCTCCCCAGCGGTATCTATATG
 GAAATGCTTACCAGAAAGCTAGTGGGTGTTCAGGAGGCTGGAGGAAGGGGGTGTGAGTGGAGAGGCG
 AGCCCACTGAGAGCTCAAACTTTCAAAAGTTGGAATGTCACTCAAGTGGCATGTGCTGTGACCAATTTATA
 TTTAGAAATTTTACRAATAGTGTCTTATCTCAAAGCAGGAATTTGGTGGCAGATTTTACAAAGATATGTA
 CCTTCCAATTTGGAATCTTCTTGTGCAATTTCTAGATAAAAGATGCGCTTGTCTATGAAATTTATA
 TAAACGATCTCTGTACAAATAATGTATTCAATACCAATACAGATCTTGAATGTCTTCCCTTACTAC
 50 TTTTGTGTCCCAAGTATATATCTGAAGTTTATTTTATTTTAGTGTCTGAGGTT (SEQ ID NO:6695)

gi|179982|gb|M57729.1|HUMCCC5 Human complement component C5 mRNA, complete cds
 CTACCTCCAACTGTGGGCTTTTGGGAATCTTGTGTTTATCTTCTGGGGAAGAACCTGGGAGCAGG
 55 AGCAACATATGTCACTTTTCAGCACCAAAAATATTCGCTGTGGAGCATCTGAAAATAATTGTGATTTCAAGT
 TTAGGATACACTGAAGCATTGTGATGCAACAATCTCTATTAAGATTTATCTGATATAAAATTTAGTTAC

TCCTCAGGCCATGTTCAATTTAATCTCAGAGAATAAATCCAAAACCTCTGCAATCTTAAACAATACAACCAA
 AACATTTGCCCTGGAGGACAAAACCCAGTTCTTCTATGTGATTTGGGAAGTTGTATCAAAAGCATTTTTCAAA
 ATCAAAAAGAAATGCCAATAACCTATGACAAATGGATTTCTTCAATCATACAGACAAAACCTGTTTATATCT
 CCAGACCCAGTCAGTAAAGATTAGAGTTTATCTCGTTGAATGACGACTTGAAGCCAGCCAAAAGAGAACTG
 5 TCTTTAACTCTTCATGATCTCTGAAGAGTCAGAAATTTGACATGGTGAAGAAATGTATCATATTTGGAATTTAT
 CTCTTTTCTGACTCTCAAGATTCCCTCTAATCTAGATATGGTATGTTGGACGATCAAGGCTAAATATAAA
 GAGGACTTTTCAACACTGGAAGCCGATATTTTGAAGTTAAAGAAATATGCTCTTGGCCACATTTTCTGTCT
 CATCTCAGCCAGAAATATAATTTCAATGGTTACAAGAACTTTTGAATTTTACTATAAAGCAG
 10 ATATTTTATAAATAAGTAGTCACCTGAGGCTGACGTTTATATACATTTTGGAAATAGAGAAAGACTTAA
 GATGATCAAAAGAAATGACGAAAACAGCAATGCAAAAACAAATGTTGATAATTTGGAAATGCTCAAGTCA
 CATTTGATTTTGAACAGCAGTCAAAAGACTGTCACTACAGATTTAGAAGTTTAAACAAACAGTACCT
 TTATATTGCTGTAAACAGTCATAGAGTCTACAGGTGGATTTTCTGAAGAGGCAAGAAATCCTGGCATCAAA
 TATGTCTCTCTCCCTCAAACTGAATTTGGTTGCTACTCTCTTTTCTGGAAGCTGGGATTTCCATATCT
 15 CCATCAAGGTCAGGTTAAAGATTGCTTGACCACTTGGAGTGGTAGGAGGAGTCCCAAGTAACTGATTAAGTCA
 AACAAATGTGTAAACCAAGACATCTGACTTGGATCCAAGCAAAAGTGTAAACAGCTGTGATGATGGA
 GTAGCTTCTCTTGTGCTTAATCTCCATCTGGAGTGCAGGTGCTGGAGTTTAAATGTCAAAATGATGCTC
 CTAGCTCTTCCAGAGAAATCAGGCCAGGGAAGTTACCGAGCAATAGCATCTCATCTCTCAGCCAAAG
 20 TTACTTTTATTTGATTTGGACTGATAACCAATAAGCTTTTGTAGTGGAGAACATCTGAATATTTATTTGTT
 ACCCCCAAAGGCCCATATATTGACAAAATACTCACTATAATTAATCTGATTTTATCTCAAGGGAACAAATTA
 TCCATTTTGGCCAGGAGGAGAAATTTTCAGATGCATTTATCAAAAGTATAAACTCTCCAGTAACAGAGAA
 CATGGTTCTCTCATCCGACTTCTGGTCTATTATATCTGACAGGAGAACAGACAGCAGAAATTAGTGTCT
 GATTCAGCTCGGTTAAAGATTGAAGAAATAATGGGCAACAGCTCCAGGTTTCATCTGCTCTCTGATGACG
 25 ATGCATATTTCTCAAGCCAACTGTGCTCTTAATAATGGCACTGGAATGGATTTCTCGGGTGGCATTAGC
 AGCAGTGGACAGTGCTGTGATGAGTCTCAAAGAGGAGCCAAAAGGCCCTTGGAAAGAGTATTTCAATCT
 TTAGAGAGAGTGATCTGGGCTGTGGGCGAGGTGGTGGCTCTCAACAATGCCAATGTGTTCTACCTAGCTG
 GACTTACTCTTCTCAACTAATGCAAAATGCAGATGACTCCCAAGAAATGATGAACCTTTGAAAGAAATTTCT
 CAGGCCAAGAGAACCTGTCAAAAAGAGATAGAAGAAATAGCTGCTAAATATAAACATCTCAGTATGAGG
 30 AAATGTGTACGATGAGCTCGCTTAATAATGATGAACCTGTGAGCAGCAGCTGCACGAGTAGTT
 TAGGGCCAGATGCACTCAAGCTTTCTACTGAATGTTGTGCTCGCAAGCCAGCTCCGCTGCTAATATCTC
 TCATAAAGACATGCAATTTGGGAAGCTACACATGAAGACCTGTTACCAAGTAGCAAGCCAGAAATCTGGC
 AGTTAATTTCCAGAAAGCTGGTTGTGGGAAGTTTCATCTTGTTCAGAGAGAAACAGTTCAGTTTCCGCC
 35 TACTGATTTCTTAACACCTTGGGAAATTCAGGCAATTTGGCAATTTCAACACTGGATATATGTTGTCTGTA
 TACTGTCAAGGCAAGGTGTTTCAAAGAGTCTTCTGGAAATGAATATACCATATTTCTGTTGTACAGGGA
 GAACAGATCCAAATGAAAGGAACGTGTTTACAACATAGGACTTCTGGGATGCAGTTCTGTGTTTAAATGT
 TCTGCTGTGGAGGGAATCTGCACCTTCGGAAGCCAGTCAATGATCATCAGGTCACAGGCTCTTCAATGT
 40 TGTGCGCCAGAAAGTAGAGGGCTCCTTCAGTCACTTGTGTGACATTCACTGTGCTTCTCTGGAAATTTGCG
 CTTCACACATCAATTTTCACTGGAGACTTGGTTTGGAAAGAAATCTTAGTAAAAACATTACGAGTGG
 TGCCAGAGGTTGTCAAAAGGGAAGCTATTCTGGTGTATCTTTGGATCTAGGGGATTTATGTTACCAT
 TAGCAGACGAAAGGAGTCTCCATACAGGATACCTTTAGATTTGGTCCCAAAACAGAAATCAAAAGGATT
 45 TTGAGTGTAAAGAGACTGCTGTAGGTGAGATCTGTCTGAGTTCTAAGTCAGAGCAAGTAAATAAATCTC
 TAACCCACTCTCCCAAGGAGTGCAGAGGCGGAGCTGATGAGCTTGTCCAGTATCTATGTTTTCATA
 CTACCTGGAAACAGGAAATCATTTGGAACATTTTTCATCTGACCCATTAAATTTGAAGAGCAGAACTGAG
 AAAAAATTTAAAGAGGGAATGTTGAGCATATGCTCTACAGAAATGCTGACTACTCTTACAGTGTGTGGA
 50 AGGCTGTGAAGTCTAGCACTTGGTTAAACAGCTTTTGTCTTAAAGATCTTGGCAAGTAAATAAATGAT
 AGAGCAGAACCAAAATTCATTTGTAATTTCTTTATTTGGGCTAGTTGAGAATTAATCAATTAGATAATGGA
 TCTTTCAAGGAAATTCAGAGTATCAACCAATAAAATTAACGGGTACCTTGCCTGTGTAAGGCCAGAGGA
 ACAGCTTATATCTTACAGCCTTTACTGTGATTTGGAATTAGAAAGGCTTTGCGATATATGCCCTTGGTGA
 55 AATTCGACACAGCTCTAATTAAGGCTGACAACTTTCTGCTTGAATTAAGCTCAGGCCAGAGCAGTAAAT
 ACATTTGGCCATTTCTGCGTATGCTCTTTTCCCTGGGAGATTAAGCTCACCACAGTTCGCTCAATTTGTT
 GAGCTTTGAAGAGAGAGCTTTGGTTAAAGGTAATCAACCATTTATCTGTTTGAAGAACCAATCTTCA
 GCATAAAGACAGCTCTGTGACTTCAACACTGGTACGGCAGTGTGTAAGAACCACTGCTATGCTTTTACTC
 ACCAGCTCTGAATCTGAAAGATATAAATATGTTTAAACCCAGTCATCAATGGCTATCAGAGAGCAGAGGT
 ATGGAGGTGGCTTTTATTCACCCAGGACACCATCAATGGCATTTGAGGGCTCTGAGGAATATTCATCTCT
 GGTATAACACTCTCGCTGAGTATGGACATCGATGTTTCTTACAAGCAATAAAGGTGCTTTACATTAATAT
 60 AATTTGACAGACAGAAATTTCTTGGGAGGCGAGTAGAGGTGCTCTCAATGATGACCTCATGTGTGATGTA
 CAGAAATTTGGCAGTGGCTTGGCTACAGTACATGTAAACACTGTAGTTTCAAAAACAGTACCTCTCGAGGA
 ATGTTTGCAGCTTTGATTTGAAATCGATACCTCAGGATATTGAGCATCCCATACAGAGGCTACGGAAC
 TCTGATTAACAAACGATAGTAGCATGTGTCAGCTACAAGCCAGCAGGGAAGATCATCATCTGATGCTT

356

5

gi|7110682|ref|NM_011044.1| Mus musculus phosphoenolpyruvate carboxykinase 1,
45 cytosolic (Pck1), mRNA

357

5 CCTATGTGGTGGCCAGCATGCGGATCATGACTCGGATGGGCATATCTGTGCTGGAGGCCCTGGGAGATGG
GGAGTTCATCAAGTGCTGCACCTCTGTGGGGTGCCCTCTCCCTTAAAAAAGCCTTTGGTCAACAACATGG
10 GCTTGCAACCCCTGAGCTGACCTTGATCGCCCACTCCCGGACCGCAGAGAGATCATCTCCTTTGGAAGCG
GATATGGTGGGAACCTACTACTCGGGAAGAAATGCTTTGCGTTGCGGATGCCAGCCCTTGGCTAAGGA
GGAAAGGTGGCTGGCGGAGCATATGCTGATCTCTGGGCATAACTAACCCGAAAGCAAGAAGAAATACCTG
15 GCCCGAGCCCTTCCTTAGTGCTGTGGGAAGCATAACTGGCCATGATGAACCCAGCCCTGCCCGGGTGGGA
AGGTTCGAATGTGTGGGCGATGACATTCCTCTGATGAAGTTTGATGCCCAAGGCAACTTAAGGGCTATCAA
CCCAGAAAAACGGGTTTTTTGGAGTTGCTCTCTGGCACCTCAGTGAAGACAATCCAAATGGCCATTAAAAAC
20 ATCCAGAAAAACACCATCTTCACCAACGTGGCCGAGACTAGCGATGGGGGTGTTTACTGGGAAGGCATCG
ATGAGCCCGTGGCCCCGGGAGTCACCATCACCTCTGGAAGAACAAAGGAGTGGAGACCGCAGGACGCGGA
ACCATGTGCCCATCCCACTCGAGATTCTGCACCCCTGCCAGCCAGTGCCCCATTATTGACCTTGCCCTGG
GAATCTCCAGAAAGGAGTACCCATTGAGGGTATCATCTTTGGTGGCCGTAGACCTGAAGGTGTCCCCCTTG
25 TCTATGAAGCCCTCAGCTGGCAGCATGGGGTGTTGTAGGAGCAGCCATGAGATCTGAGGCCACACGCTGC
TGCAAGAACACAAGGGCAAGATCATATGCACGACCCCTTTGCCATGGACCCCTTCTCGGCTACAACTTC
GGCAAACTCCTGGCCCATGGCTGAGCATGGCCCAACGCCAGCAGCCAAAGTTGCCCAAGATCTTCCATG
TCAAATGGTTCGGCAAGGACAAAGATGGCAAGTTCTCTGGCCAGGCTTTGGCGAGAACTCCCGGGTGTCT
30 GGAAGTGGATGTTGGGCGGATTTGAAGGGGAAGACAGCGCCAAAGCTCACGCCCATCGGCTACATCCCTAAG
GAAAACGCTTTGAACTCGAAAGGCTGGGGGGCGTCAACGTGGAGGAGCTGTTTGGGATCTCTAAGGAGT
TCTGGGAGAAGGAGGTGGAGGAGATGACAGGTATCTGAGGAGCAGGCTCAACACCGACCTCCCTTACGA
25 AATTGAGAGGGAGCTCCGAGCCCTGAAACAGAGAACTAGCCAGATGTAAATCCCAATGGGGGCGTCTCGA
GAGTCACCCCTTCCCACTCACAGCATCGCTGAGATCTAGGAGAAAGCCAGCTGTCTCAGCTTTGAGATA
GCGGCACAATCGTGAGTAGATCAGAAAAAGCACCTTTTAATAGTCAGTTGAGTAGCACAGAAACAGGCTA
30 GGGGCAAAATAAGATTGGGAGGGGAATACCCGCATAGTCTCTGAAAGTTTGCAATTGACACCAATGGGGGT
TTTGGTTCCACTTCAAGGTCACTCAGGAATCCAGTTCTTCACGTTAGCTGTAGCAGTTAGCTAAAAATGCA
CAGAAAAACATACTTGAGCTGTATATATGTGTGTGAACGTGTCTGTGTGAGCATGTGTGTGTGTGTGT
TGT
AACCTTTGGGGAAAAATCTTGGGCAAAATTTGTAGCTGTAACTAGAGAGTCATGTGTCTTTGTGTGTGTGT
TGTATGTTTAAATTAATTTTATACACCGCCCTTACCTTTCTTACATAAATTGAATTTGGTATCCGAGCCA
CTTCTTGGGAAAAAATTACAAAAATAA (SEQ ID NO: 6699)

30

Example 5. siRNAs decrease mRNA levels *in vivo*

Male CMV-Luc mice (8-10 weeks old) from Xenogen (Cranbury, NJ) were administered cholesterol conjugated siRNA (see Table 17).

Table 17. Solutions administered to mice

Group	N	Injection Mix
1	7	Buffer (PBS [pH 7.4])
2	8	Cholesterol conjugated siRNA (ALN-3001)

Table 18. Test siRNA agents targeting Luciferase

siRNA	Sequence
ALN-1070	5'-GAA CUG UGU GUG AGA GGU CCU-3' (SEQ ID NO:5277) 3'-CG CUU GAC ACA CAC UCU CCA GGA-5' (SEQ ID NO:5278)
ALN-1000	5'-GAA CUG UGU GUG AGA GGU CCU-GS-3' (SEQ ID NO:5279) 3'-CG CUU GAC ACA CAC UCU CCA GGA-5' (SEQ ID NO:5280)
ALN-3000	5'-GAA CUG UGU GUG AGA GGU CCU-3' (SEQ ID NO:5281) 3'-CS ¹ GS ¹ CUU GAC ACA CAC UCU CCA GGA-5' (SEQ ID NO:5282)
ALN-3001	5'-GAA CUG UGU GUG AGA GGU CCU-cho1. ² -3' (SEQ ID NO:5283) 3'-CS ¹ GS ¹ CUU GAC ACA CAC UCU CCA GGA-5' (SEQ ID NO:5284)

¹ 2' O-Me group is attached to the nucleotide and the nucleotides have phosphorothioate linkages (indicated by "s")

2 cholesterol is conjugated to the antisense strand via the linker: U-pyrroline carrier-C(O)-(CH₂)₇-NHC(O)-cholesterol (via cholesterol C-3 hydroxyl).

Animals were injected (tail vein) with a volume of 200-250 μ l test solution containing buffer or an siRNA solution. Group 1 received buffer and group 2 received cholesterol conjugated siRNA (ALN-3001) at a dose of 50 mg/kg body weight. Twenty-two hours after injection, animals were sacrificed and livers collected. Organs were snap frozen on dry ice, then pulverized in a mortar and pestle.

For Luciferase mRNA analysis (by the QuantiGene Assay (Genospectra, Inc.; Fremont, CA)), approximately 10 mg of tissue powder was resuspended in tissue lysis buffer, and processed according to the manufacturer's protocol. Samples of the lysate were hybridized with probes specific for Luciferase or GAPDH (designed using ProbeDesigner software (Genospectra, Inc., Fremont, CA) in triplicate, and processed for luminometric analysis. Values for Luciferase were normalized to GAPDH. Mean values were plotted with error bars corresponding to the standard deviation of the Luciferase measurements.

Results indicated that the level of luciferase RNA in animals injected with cholesterol conjugated siRNA was reduced by about 70% as compared to animals injected with buffer (see FIGs. 8A and 8B).

In Vitro Activity

HeLa cells expressing luciferase were transfected with each of the siRNAs listed in Table 18. ALN-1000 siRNAs were most effective at decreasing luciferase mRNA levels (~0.6 nM siRNA decreased mRNA levels to about ~65% the original expression level, and 1.0 nM siRNA decreased levels to about ~20% the original expression level); ALN-3001 siRNAs were least effective (~0.6 nM siRNA had a negligible effect on mRNA levels, and 1.0 nM siRNA decreased levels to about ~40% the original expression level).

Pharmacokinetics/Biodistribution

Pharmacokinetic analyses were performed in mice and rats. Test siRNA molecules were radioactively labeled with ^{32}P on the antisense strand by splint ligation. Labeled siRNAs (50mg/kg) were administered by tail vein injection, and plasma levels of siRNA were measured periodically over 24 hrs by scintillation counting. Cholesterol conjugated siRNA (ALN-3001) was discovered to circulate in mouse plasma for a longer period of time than unconjugated siRNA (ALN-3000) (FIG. 9). RNase protection assays indicated that cholesterol-conjugated siRNA (ALN-3001) was detectable in mouse plasma 12 hours after injection, whereas unconjugated siRNA (ALN-3000) was not detectable in mouse plasma within two hours after injection. Similar results were observed in rats.

Mouse liver was harvested at varying time points (ranging from 0.08-24 hours) following injection with siRNA, and siRNA localized to the liver was quantified. Over the time period tested, the amount of cholesterol-conjugated siRNA (ALN-3001) detected in the liver ranged from 14.3-3.55 percent of the total dose administered to the mouse. The amount of unconjugated siRNA (ALN-3000) detected in the liver was lower, ranging from 3.91-1.75 percent of the total dose administered (FIG. 10).

Detection of siRNA in Different Tissues

Various tissues and organs (fat, heart, kidney, liver, and spleen) were harvested from two CMV-Luc mice 22 hours following injection with 50 mg/kg ALN-3001. The antisense strand of the

siRNA was detected by RNase protection assay. The liver contained the greatest concentration of siRNA (~8-10 μg siRNA/g tissue); the spleen, heart and kidney contained lesser amounts of siRNA (~2-7 μg siRNA/g tissue); and fat tissue contained the least amount of siRNA (<~1 μg siRNA/g tissue) (FIG. 11).

5

Glucose-6-phosphatase siRNA detection by RNase Protection Assay

Balb/c mice were injected with U/U, 3'C/U, or 3' C/3' C siRNA (4 mg/kg) targeting glucose-6-phosphatase (G6Pase) (see Table 19). Administration was by hydrodynamic tail vein injection (hd) or non-hydrodynamic tail vein injection (iv), and siRNA was subsequently detected in the liver by RNase protection assay.

10

Table 19. Test siRNA agents targeting glucose-6-phosphatase

siRNA	Description
U/U	No cholesterol; dinucleotide 3' overhangs on sense and antisense strands
3'C/U	dinucleotide 3' overhangs on sense and antisense strands; cholesterol conjugated to 3' end of sense strand (mono-conjugate)
3'C/3'C	dinucleotide 3' overhangs on sense and antisense strands; cholesterol conjugated to 3' end of both sense and antisense strands (bis-conjugate)

15

Unconjugated siRNA (U/U) delivered by hd was detected by 15 min. post-injection (the earliest determined time-point) and was still detectable in the liver 18 hours post-injection (FIG. 12).

20

Delivery by normal iv administration resulted in the greatest concentration of 3'C/3'C siRNA (the bis-cholesterol-conjugate) in the liver 1 hour post injection (as compared to the mono-cholesterol-conjugate 3'C/3'U siRNA). At 18 hours post injection, 3'C/3'C siRNAs and 3'C/U siRNA were still detectable in the liver with the bis-conjugate at higher levels compared to the mono-conjugate (FIG. 13).

Example 6. siRNAs decrease protein activity levels *in vivo*

25

Male CMV-Luc mice were bred by Charles River Laboratories, Inc. (Wilmington, MA). Mice (6-7 weeks old) were administered cholesterol conjugated siRNA (see Tables 20-22).

Table 20. Test groups for *in vivo* siRNA assays-experiment 1

Group	N	Injection Mix
1	10	Buffer (PBS [pH 7.4])
2	11	Cholesterol conjugated siRNA (ALN-3001)

Table 21. Test groups for *in vivo* siRNA assays-experiment 2

Group	N	Injection Mix
1	8	Buffer (PBS [pH 7.4])
2	8	Cholesterol conjugated siRNA (ALN-3001)

Table 22. Test groups for *in vivo* siRNA assays-experiment 3

Group	N	Injection Mix
1	8	Buffer (PBS [pH 7.4])
2	8	Cholesterol conjugated siRNA (ALN-3001)

Animals were injected (tail vein) with a volume of 200-250 μ l test solution containing buffer or an siRNA solution. Group 1 received buffer and group 2 received cholesterol conjugated siRNA (ALN-3001) at a dose of 75 mg/kg body weight. Nineteen to 22 hours after injection, animals were sacrificed and livers collected. Organs were snap frozen on dry ice, then pulverized in a mortar and pestle.

For Luciferase activity analysis, approximately 50 mg of tissue powder was resuspended in 0.5 ml Cell Lysis Buffer (Promega, Inc.). Samples were vortexed vigorously for three minutes, snap frozen in liquid nitrogen, then thawed in a 37 degree water bath. This process was repeated twice more. After the final thaw, samples were vortexed for three minutes. Insoluble material was removed by centrifugation in a microcentrifuge (4 degrees) at full speed for four minutes. Supernatants were collected. Twenty to 25 microliters of each sample were pipetted into assay tubes in triplicate, and allowed to come to room temperature. For activity measurements, a luminometer (Berthold, Inc.) was programmed to deliver 200 microliters of "Bright Glow" assay reagent (Promega, Inc.) to the test sample, and record light emission over a ten second period.

To measure total protein, samples of supernatant were diluted thirty fold, and five microliter samples were measured in triplicate in a Bradford protein microassay (Bio-Rad). Bovine Serum Albumin was used to generate a standard curve.

Luciferase activity was determined as the mean of the luminometry reading normalized to mean protein content. Mean normalized values were then calculated for the buffer and siRNA-treated groups in each experiment. For each experiment, the normalized Luc level of the siRNA treated group is expressed as a percentage of the buffer control (which was set to 100%). Error bars
5 indicate standard deviations.

Results indicated that the level of luciferase activity in animals injected with cholesterol conjugated siRNA was reduced by about 55% as compared to animals injected with buffer (see FIG. 14).

OTHER EMBODIMENTS

10 While this invention has been particularly shown and described with reference to preferred embodiments thereof, it will be understood by those skilled in the art that various changes in form and details may be made therein without departing from the scope of the invention encompassed by the appended claims.

WHAT IS CLAIMED IS:

1. A method for reducing apoB-100 levels in a subject comprising administering to a subject an iRNA agent, which targets apoB-100.
- 5 2. The method of claim 1, wherein said iRNA agent targets a sequence identical to any one of SEQ ID NOs listed in Tables 9 and 10.
3. The method of claim 1, wherein said iRNA agent comprises a cholesterol moiety.
- 10 4. The method of claim 3, wherein said cholesterol moiety is coupled to a sense strand.
5. The method of claim 3, further comprising a second cholesterol moiety.
6. The method of claim 5, wherein said second cholesterol moiety is coupled to a sense
15 strand.
7. The method of claim 1, wherein said iRNA agent is at least 21 nucleotides in length, and the duplex region of the iRNA is about 19 nucleotides in length.
- 20 8. The method of claim 1, wherein the subject is suffering from a disorder characterized by elevated or otherwise unwanted expression of apoB-100, elevated or otherwise unwanted levels of cholesterol, and/or dysregulation of lipid metabolism.
9. The method of claim 8, wherein said disorder is chosen from the group of HDL/LDL
25 cholesterol imbalance; dyslipidemias, *e.g.*, familial combined hyperlipidemia (FCHL), acquired hyperlipidemia; hypercholesterolemia; statin-resistant hypercholesterolemia; coronary artery disease (CAD) coronary heart disease (CHD) atherosclerosis
10. The method of claim 9, wherein said iRNA agent is administered to a subject suffering
30 from statin-resistant hypercholesterolemia.

11. A method for reducing glucose-6-phosphatase levels in a subject comprising administering to a subject an iRNA agent that targets glucose-6-phosphatase.

12. The method of claim 11, wherein said iRNA agent is at least 21 nucleotides in length,
5 and the duplex region of the iRNA is about 19 nucleotides in length.

13. The method of claim 12, wherein the iRNA agent is administered to a subject to inhibit hepatic glucose production, or for the treatment of glucose-metabolism-related disorders.

10 14. The method of claim 12, wherein said disorder is diabetes.

15. The method of claim 12, wherein said disorder is type-2 diabetes.

16. The method of claim 12, wherein said disorder is glitaxzone-resistant diabetes.

15

17. An iRNA agent comprising a sense sequence and an antisense sequence, wherein the sense sequence comprises one or more cholesterol moieties, and the antisense sequence targets a human gene sequence.

20 18. The iRNA agent of claim 17, wherein said human gene is an oncogene.

19. The iRNA agent of claim 17, wherein said human gene is apoB 100.

20. The iRNA agent of claim 17, wherein said human gene is glucose-6-phosphatase.

25

21. The iRNA agent of claim 17, wherein said human gene beta catenin.

22. An iRNA agent, wherein the agent targets apoB 100.

30

23. An iRNA agent, wherein the agent targets glucose-6-phosphatase.

24. An iRNA agent, wherein the agent targets beta-catenin.

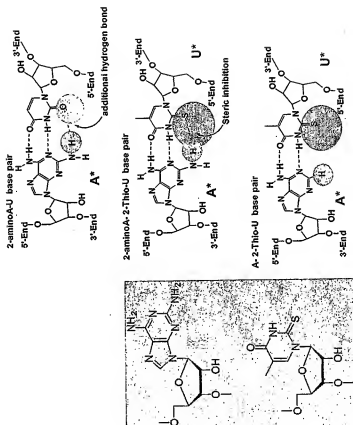
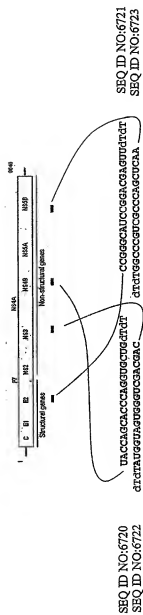


FIG. 1



3/17

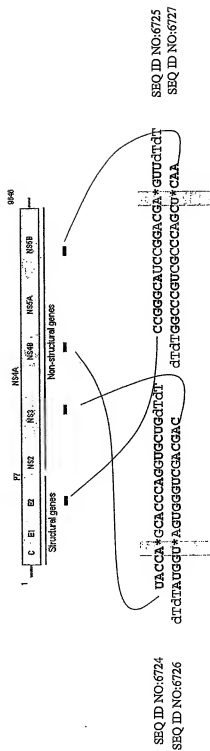


FIG. 3

4/17

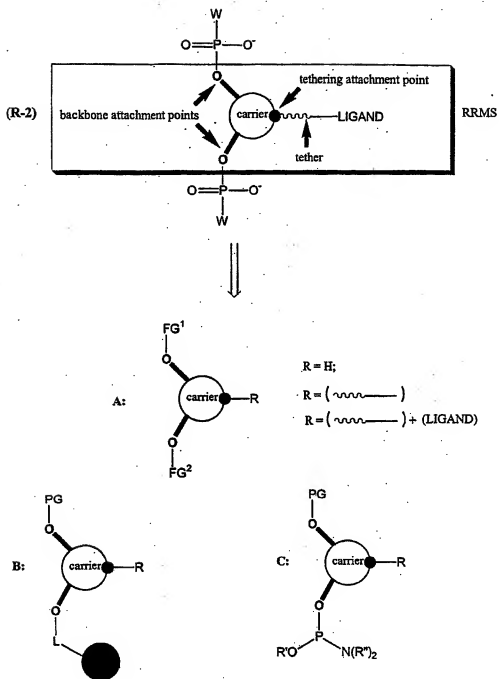


FIG. 4

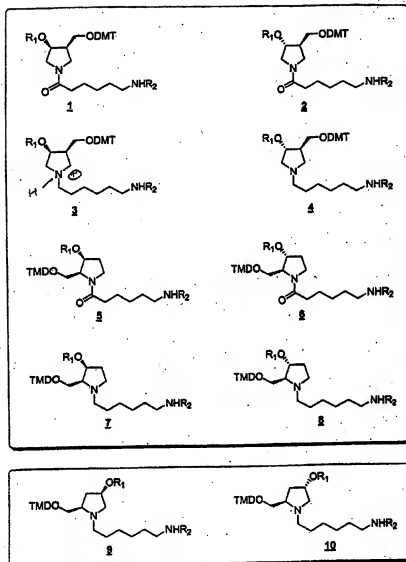


FIG. 5

6/17

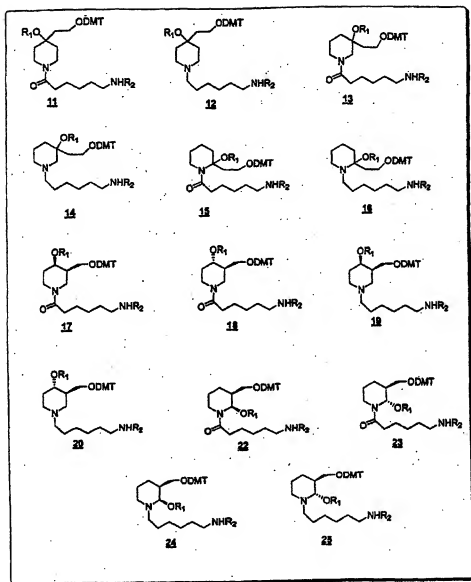


FIG. 5 (Cont'd)

7/17

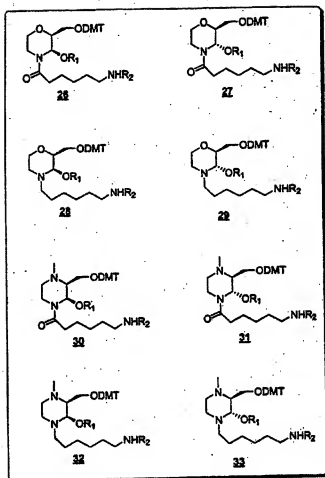


FIG. 5 (Cont'd)

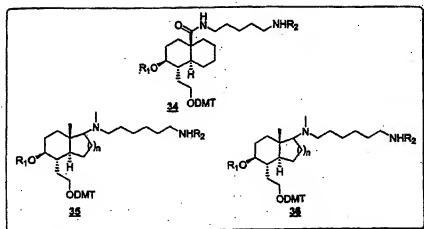


FIG. 5 (Cont'd)

9/17

FIG. 6A

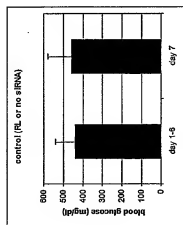


FIG. 6B

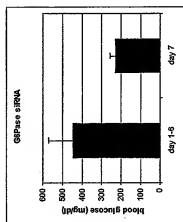
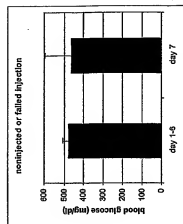
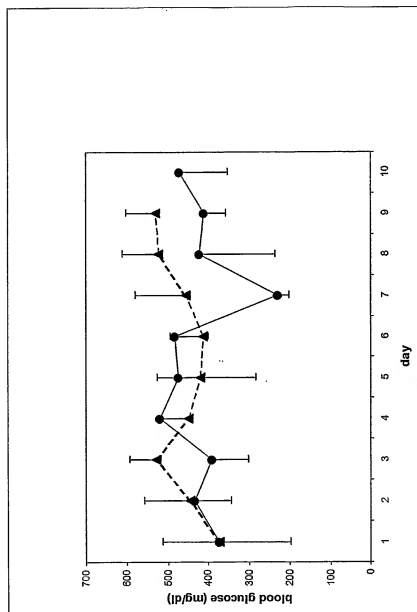


FIG. 6C



10/17

FIG. 7



11/17

FIG. 8A

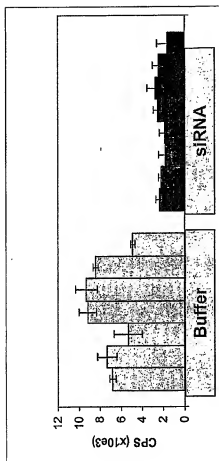
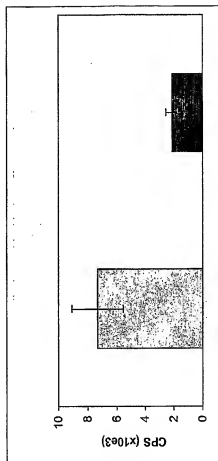


FIG. 8B



12/17

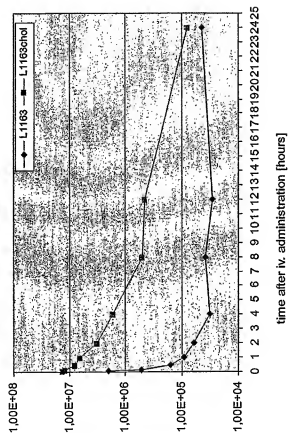


FIG. 9

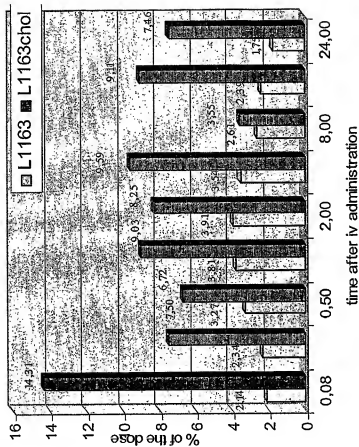


FIG. 10

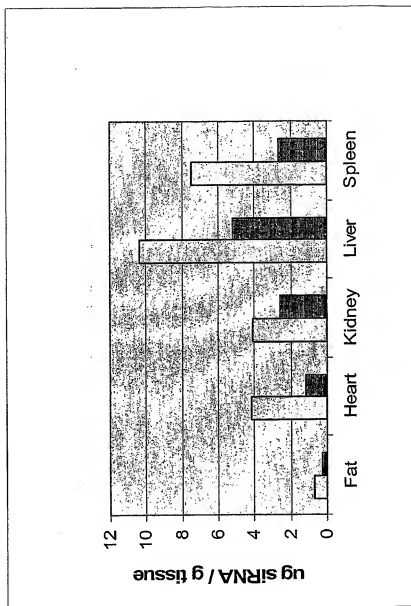


FIG. 11

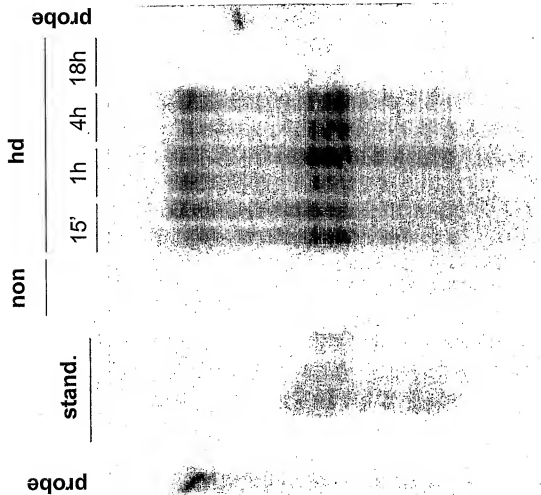
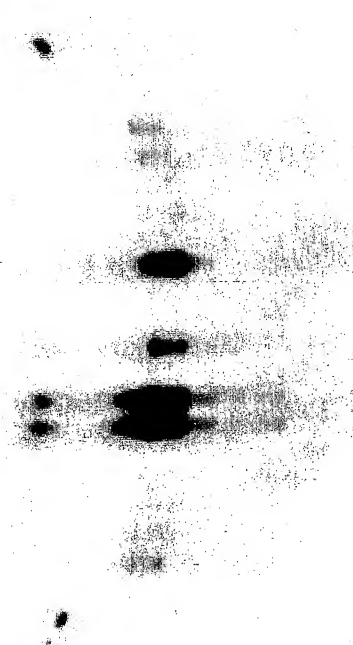


FIG. 12

16/17

probe	non		hd		iv		probe	
	stand.		u/u		3'C/3'C		3'C/u	
	1h	18h	1h	18h	1h	18h	1h	18h

FIG. 13



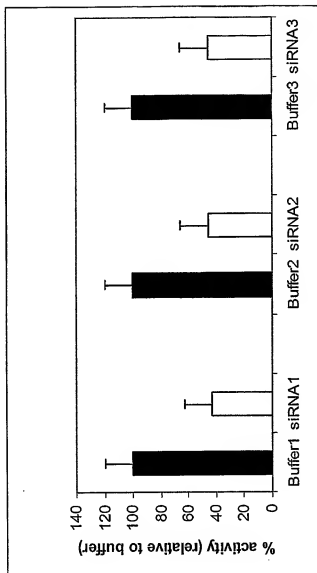


FIG. 14